Heilbronn University of Applied Sciences

Master Thesis

---

# Development and validation of a neural network for adaptive gait cycle detection from kinematic data

---

*Author:*
Alexander Kombeiz

*Student number:*
198731

*Supervisor:*
Dr. Christoph Maier

*Examiner:*
PD Dr.-Ing. Rüdiger Rupp

*A thesis submitted in fulfillment of the requirements
for the degree of Master of Science
in Medical Informatics*

January 21, 2020

# Declaration of Authorship

I, Alexander Kombeiz, declare that this thesis titled, “Development and validation of a neural network for adaptive gait cycle detection from kinematic data” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

# *Abstract*

**(1) Background:** Instrumented gait analysis is a tool for quantification of the different aspects of the locomotor system. Gait analysis technology has substantially evolved over the last decade and most modern systems provide real-time capability. The ability to calculate joint angles with low delays paves the way for new applications such as real-time movement feedback, like control of functional electrical stimulation in the rehabilitation of individuals with gait disorders. For any kind of therapeutic application, the timely determination of different gait phases such as *stance* or *swing* is crucial. Gait phases are usually estimated based on heuristics of joint angles or time points of certain gait events. Such heuristic approaches often do not work properly in people with gait disorders due to the greater variability of their pathological gait pattern. To improve the current state-of-the-art, this thesis aims to introduce a data-driven approach for real-time determination of gait phases from kinematic variables based on long short-term memory recurrent neural networks (LSTM RNNs).

**(2) Methods:** In this thesis, 56 measurements with gait data of 11 healthy subjects, 13 individuals with incomplete spinal cord injury and 10 stroke survivors with walking speeds ranging from $0.2\,\frac{m}{s}$ up to $1\,\frac{m}{s}$ were used to train the networks. Each measurement contained kinematic data from the corresponding subject walking on a treadmill for 90 seconds. Kinematic data was obtained by measuring the positions of reflective markers on body landmarks (Helen Hayes marker set) with a sample rate of 60Hz. For constructing a ground truth, gait data was annotated manually by three raters. Two approaches, direct regression of gait phases and estimation via detection of the gait events *Initial Contact* and *Final Contact* were implemented for evaluation of the performance of LSTM RNNs. For comparison of performance, the frequently cited coordinate- and velocity-based event detection approaches of Zeni et al. were used. All aspects of this thesis have been implemented within *MATLAB Version 9.6* using the *Deep Learning Toolbox*.

**(3) Results:** The mean time difference between events annotated by the three raters was $-0.07 \pm 20.17\,ms$. Correlation coefficients of inter-rater and intra-rater reliability yielded mainly *excellent* or *perfect* results. For detection of gait events, the LSTM RNN algorithm covered 97.05% of all events within a scope of $50\,ms$. The overall mean time difference between detected events and ground truth was $-11.62 \pm 7.01\,ms$. Temporal differences and deviations were consistently small over different walking speeds and gait pathologies. Mean time difference to the ground truth was $13.61 \pm 17.88\,ms$ for the coordinate-based approach of Zeni et al. and $17.18 \pm 15.67\,ms$ for the velocity-based approach. For estimation of gait phases, the gait phase was determined as a percentage. Mean squared error to the ground truth was $0.95 \pm 0.55\%$ for the proposed algorithm using event detection and $1.50 \pm 0.55\%$ for regression. For the approaches of Zeni et al., mean squared error was $2.04 \pm 1.23\%$ for the coordinate-based approach and $2.24 \pm 1.34\%$ for the velocity-based approach. Regarding mean absolute error to the ground truth, the proposed algorithm achieved a mean absolute error of $1.95 \pm 1.10\%$ using event detection and one of $7.25 \pm 1.45\%$ using regression. Mean absolute error for the coordinate-based approach of Zeni et al. was $4.08 \pm 2.51\%$ and $4.50 \pm 2.73\%$ for the velocity-based approach.

**(4) Conclusion:** The newly introduced LSTM RNN algorithm offers a high recognition rate of gait events with a small delay. Its performance outperforms several state-of-the-art gait event detection methods while offering the possibility for real-time processing and high generalization of trained gait patterns. Additionally, the proposed algorithm is easy to integrate into existing applications and contains parameters that self-adapt to individuals' gait behavior to further improve performance. In respect to gait phase estimation, the performance of the proposed algorithm using event detection is in line with current wearable state-of-the-art methods. Compared with conventional methods, performance of direct regression of gait phases is only moderate. Given the results, LSTM RNNs demonstrate feasibility regarding event detection and are applicable for many clinical and research applications. They may be not suitable for the estimation of gait phases via regression. For LSTM RNNs, it can be assumed, that with a more optimal configuration of the networks, a much higher performance is achieved.

# *Acknowledgements*

I would like to thank my supervisor, Dr. Christoph Maier, for providing me with expert advice and feedback in course of this thesis, as well as his engagement through the learning process of this thesis.

Also, I want to thank my examiner, PD Dr.-Ing. Rüdiger Rupp. Without the collaboration with the *Section Experimental Neurorehabilitation of the Spinal Cord Injury Center of Heidelberg University Hospital*, this thesis would not have been possible. PD Dr.-Ing. Rüdiger Rupp familiarized me with the task and provided me with the equipment to handle this thesis. I thank him for his extraordinary support, his helpful advice and expertise and his willingness to help.

Special thanks go to my colleague Christian Schuld, who took care of me during this thesis. He was always willing to help and supported me with approaches, guidelines and feedback. He introduced me to the topic and supported me through the way of it. Whenever I got off the path, he brought me back to it. I express my gratitude for his useful comments and remarks during the fine-tuning of this thesis. Also, I want to thank him for his help during the manual annotation, especially for his self-developed script for manual annotation of gait data, which saved a lot of time. I am very grateful for having him as my colleague.

Finally, I thank my colleague Patrick Jersch for his support and his help during the manual annotation.

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **ANN** | Artificial neural network |
| **BA** | Bland-Altman |
| **CS** | Complete set |
| **ED-FNN** | Exponentially delayed fully connected neural network |
| **FC** | Final contact |
| **FN** | False negatives |
| **FP** | False positives |
| **FVA** | Foot velocity algorithm |
| **HMA** | Hreljac-Marshall algorithm |
| **HPA** | High pass algorithm |
| **HTR** | Hierarchical translation-rotation |
| **IC** | Initial contact |
| **ICC** | Intraclass correlation coefficient |
| **IMU** | Inertial measurement unit |
| **iSCI** | Incomplete spinal cord injury |
| **iSCI H** | Incomplete spinal cord injury with more impairment |
| **iSCI L** | Incomplete spinal cord injury with less impairment |
| **LOA** | Limits of agreement |
| **LSTM** | Long short term memory |
| **MAE** | Mean absolute error |
| **MIMU** | Magnetic and inertial measurement unit |
| **MS** | Multiple sclerosis |
| **MSE** | Mean squared error |
| **RNN** | Recurrent neural network |
| **TN** | True negatives |
| **TP** | True positives |
| **TRB** | Three-dimensional trajectory |

# Contents

# 1 Introduction

Walking is a fundamental human ability that allows human beings to independently pursue their everyday activities and be productive members of society. [93] Already slight changes of the physiological walking pattern do not only lead to musculoskeletal pain, which restricts independence in everyday life, but are also linked to a higher risk of falling and may contribute to cognitive impairments. [2]
To reveal information related to health and well-being, gait analysis is used as a tool to quantify a human's gait function. [93, 2] The quantification of gait enables a comparison between different subjects and conditions and can be used in a wide range of applications including the evaluation of neurological diseases, orthopedic disabilities or progress during rehabilitation. Additionally, it can be used in real-time feedback applications, like prosthetics, gait re-training or control of functional electrical stimulation. [17, 47, 24, 51, 92]
In such applications, the gait cycle is interpreted as a sequence of consecutive gait phases, mainly characterized as the stance and the swing phase. Depending on the application, a more sequencing breakdown of the gait cycle might be required. Up to eight different gait phases have been considered. [47, 82]
For recording gait data, different sensor systems can be used. These systems can be subdivided into wearable systems, like foot pressure insoles, footswitches, accelerometers and gyroscopes, and non-wearable systems, such as optical marker-based systems or force platforms. While some systems have proven to be superior to others, there is no optimal solution. Fundamental problems, and thus topics of active research, are selection of proper sensor and computational methodology. [82]

## 1.1 Thesis Contributions

The main contribution of this thesis is the development and validation of a neural network algorithm for real-time detection of gait events and real-time regression of gait phases in subjects with physiological or pathological gait patterns.
A secondary contribution is the validation of the current gait event detection algorithm used in the optical marker-based system of the *Spinal Cord Injury Center of Heidelberg University Hospital*.

### Background

Basis of this thesis is the research project *SensSCI*, conducted at the *Spinal Cord Injury Center of Heidelberg University Hospital*. The aim of this research project is the implementation of an individualized real-time haptic sensory feedback system and its evaluation in subjects with incomplete spinal cord injury (iSCI). Through gait-phase-matched stimulation of spared afferents of sensory functions in the lower limbs, it aims to enable individuals to substantially normalize their gait pattern and thereby recovering their ambulatory function. The exact timing of the provided feedback is determined by real-time analysis and recording of gait data with an optical marker-based gait analysis

system.
The concept of gait normalization through adaptive biofeedback was proofed successfully in a previously conducted proof-of-concept study [75] by the *Spinal Cord Injury Center*. This study investigated, if subjects with gait pathologies were able to normalize their gait pattern by verbal feedback given from an implemented technical feedback system called *RehaGait*. The participants succeeded in normalizing their given feedback parameters significantly, thus this study provides a basis for using real-time feedback as part of a gait training program.
To achieve a therapeutic effect, sensory stimulation must be applied at a specific point of the gait cycle. For determination of this time instance, an optimized gait phase detection algorithm is indispensable. In own previous work, the author already developed a system for processing gait data using the *RehaGait* data set. The system consists of an algorithm for detecting gait events using a threshold of foot and heel marker trajectories in the sagittal plane and a post-processing algorithm for predicting gait phases based on time-distances between detected events. In comparison with the currently used algorithm for detection of gait events in the *Spinal Cord Injury Center*, the developed algorithm proved to be superior. However, due to use of a flawed ground truth during development, only a moderate performance was achieved in the application environment. The algorithm is therefore not optimal for use in the feedback system.
For this reason, a new approach for the determination of gait phases was developed in this thesis. With the intention to use the finished algorithm in the feedback system of *SensSCI*, following additional requirements need to be met:

- The algorithm should be able to be integrated in existing system with as few changes as possible
- The algorithm should have a considerably high recognition rate of gait events with as little delay as possible in order to allow for real-time gait feedback
- The algorithm should adapt automatically to individuals' gait behaviors in order to achieve the highest possible recognition rate of gait events

**Algorithm Framework**

This thesis provides an algorithm for detection of gait events and regression of gait phases in real-time. Since in gait analysis there is no optimal solution for these cases of application, this thesis aims to make a further contribution in order to come closer to one. The algorithm is intended for use in optical marker-based systems in applications, where real-time feedback is needed. As mentioned before, such applications may be gait re-training or functional electrical stimulation [17, 47, 75]. Mainly, a reliable use in the vibratory feedback system of *SensSCI* shall be achieved.
In conventional approaches [17, 25, 38, 56, 84], the data processing consists often of rather simple heuristics of coordinates or velocities. Such approaches do not generalize well to pathological gait due to variability in kinematics and anatomy of subjects, as well as presence of assistive devices [51]. In contrast, this thesis focuses on a data-driven processing and considers the stride as an interdependent sequence. By considering the history of past values in each processing step, a better performance and real-time processing shall be achieved. In addition, two cases of applications, detection of gait events and regression of gait phases, are combined in this algorithm in order to determine, which one is better suited for determining gait phases. In order to achieve a high granularity, gait phases are determined as percentage values.
The proposed algorithm contains four separate long short-term memory recurrent neural

networks (LSTM RNNs) [32], each of which fulfills one case of application on the corresponding side of the body. So far, this approach was only adopted in a similar way by Kidziński et al. [51].

**Validation of the current algorithm used in the *SCI Center***

For development of the proposed algorithm, gait data recorded with the optical marker-based gait analysis system of the *Spinal Cord Injury Center of Heidelberg University Hospital* is used. In order to detect gait events in recorded data, the *Spinal Cord Injury Center* uses a self-developed algorithm. However, the development took place several years ago. To evaluate the actuality of the self-developed algorithm, its performance is compared with the proposed algorithm on a common ground truth.

## 1.2 Thesis Outline

The following parts of this thesis are organized as follows:

Chapter 2 provides background information about human gait, the granularity of gait phases and the axes of human movement. Furthermore, sensors for recording gait data, an introduction for neural networks, and existing methods for detection of gait events and estimation of gait phases are presented.

Chapter 3 describes the proposed algorithm and the investigated data set. This includes subjects, experimental protocol, data annotation and feature selection as well as the development process, the architecture of the algorithm and the evaluation procedure with the used metrics.

Chapter 4 presents the results of manual annotation and performance evaluation. Results include inter-rater and intra-rater reliability, as well as the comparison of performance in chosen evaluation metrics between the proposed algorithm and other selected methods regarding event detection and phase estimation.

Chapter 5 compares the results of Chapter 4 with existing methods to discuss strengths and limitations of the proposed algorithm. Design decisions during development are reflected and new approaches for optimization are suggested.

Chapter 6 summarizes the new insights of the discussion. A final conclusion is given on the proposed algorithm regarding strengths and weaknesses, as well as an outlook for future works.

# 2 Related Works

This chapter provides necessary background information for the understanding of this thesis. Existing methods for detection of gait events and estimation of gait phases are presented.

## 2.1 Background

A brief overview of human gait, human movement axes and methods for recording human gait, as well as an overview of the neural networks used in the proposed algorithm is given in this section.

### 2.1.1 Human Gait

Human gait is a complex cyclical process consisting of the synergy of muscles, skeletal structures and the nervous system. The purpose of gait is to support the upright position and maintaining balance during static and dynamic conditions in order to move the body safely and effectively. [10] While walking is the most basic way of human displacement, it is also one of the most difficult maneuvers the body can perform. [82, 10]. Minor alteration to physiological gait patterns can increase energy expenditure and may produce an abnormal gait, while severe abnormalities can even lead to an inability to walk. [26, 89]

Under normal physiological conditions, human gait consists of a repetition of the same basic pattern. The investigation of this basic pattern can be used as a measure for quantifying functionality of the locomotor system. [73] For a quantitative comparison, gait data is aligned to predefined landmarks in the gait cycle. These landmarks are referred to as gait events and allow a segmentation of the gait into comparable phases. Two well-established parameters of alignment are the foot-contact event and the foot-off event. For generalization, these events can be also referred to as *Initial Contact* (IC) and *Final Contact* (FC). In physiological gait, these are equivalent to *Heel strike* and *Toe off.* [51] Commonly, gait phases are estimated using heuristics of recorded gait data like marker trajectories or joint angles [47, 86] or via time intervals between two successive occurrences of the same gait event [3, 72, 89]. In real-time applications, it is necessary to evaluate whether the heuristic approach can be adapted to online processing. For the real-time estimation via time intervals between detected gait events, times of future events are predicted based on the characteristics of already detected events.

For gait phases, mainly the timing information of start and duration are of importance. Among other parameters, like stride length and gait speed, the timing information is used to quantify the gait. The phase itself can be expressed as a percentage between the values 0 and 1 [47, 76], or as a class such as *stance* or *swing* [86, 92]. An established assumption is the division of the gait cycle into a stance phase of 60 percent of the entire cycle and a swing phase of 40 percent. Two additional periods of double-limb support exist, when both left and right lower extremities contact the ground in opposite synchronization of *Heel Strike* and *Toe off.* [26] An example of a stride with four different gait phases is shown in Figure 2.1. The *Heel strike* or IC marks the beginning of the stance phase while *Toe off* or FC marks the termination of the stance and beginning of the swing phase. Considering the gait events of the other leg, the phases of double-limb support can be added to the cycle at each transition between stance and swing phase. [69] Accordingly, the two individual stance phases can also be referred to as phases of single-limb support. Over time, several partitioning models with up to eight levels of granularity have been proposed. In case of more gait phases, more gait events are involved. However, there is no standard for naming the various granularities of gait phases. [82]

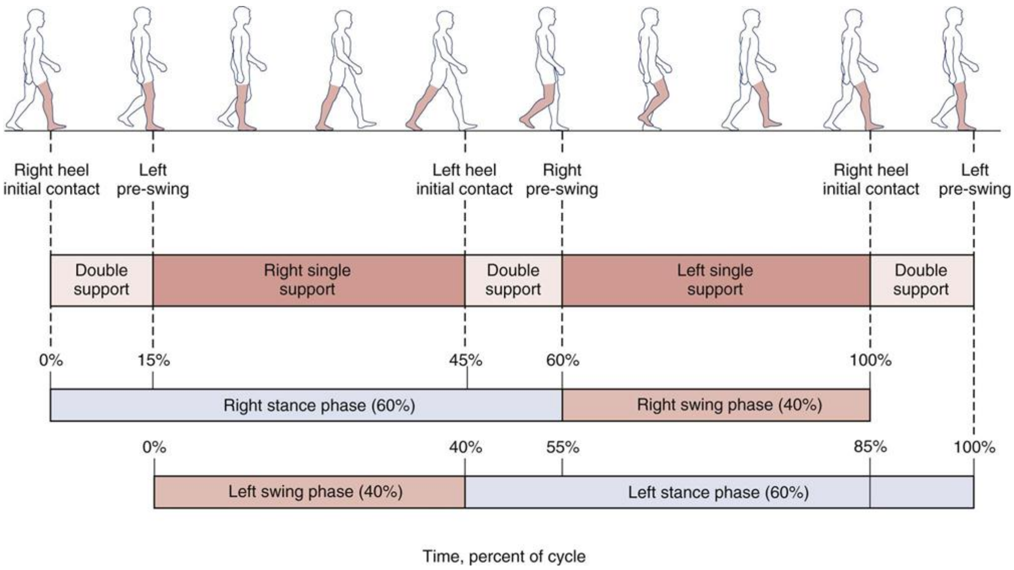


Figure 2.1: Time dimensions of the gait cycle [45]

### 2.1.2 Movement of the Human Body in Space

The movement of the human body is divided in three different planes and six different directions. The individual planes are:

- Sagittal plane, which divides the body into right and left halves
- Frontal plane, which divides the body into front and back halves
- Transverse plane, which divides the body into upper and lower halves

The placement of the planes and the possible directions of movement are shown in Figure 2.2. This distinction into different planes and directions is necessary to make an unique identification about the movement of body segments and thus define the inertial parameters. [10] In the following parts of this thesis, movements in the presented directions are abbreviated as shown in Table 2.1.

Table 2.1: Axes of abbreviation for the six fundamental directions

| Abbreviation | Positive direction of movement | Negative direction of movement |
|---|---|---|
| X | anterior | posterior |
| Y | left / contra-lateral | right / lateral |
| Z | superior | inferior |

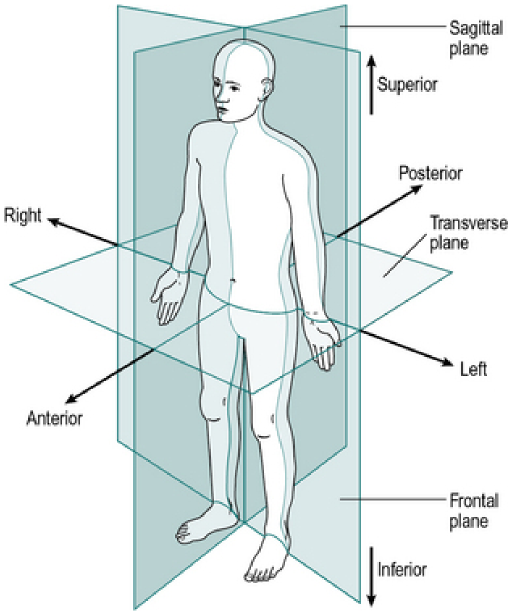


FIGURE 2.2: The anatomical position of the human body with the three reference planes and six fundamental directions [5]

### 2.1.3 Sensors

The technical equipment used to extract information from human gait can be classified into two different types: wearable sensors and non-wearable sensors. While wearable sensors are smaller compared to non-wearable sensors, and therefore used more frequently to collect gait data in outdoor environment, non-wearable sensors are mainly used in indoor research facilities under controlled conditions. Additionally, non-wearable sensors are based on costly equipment and need a complicated setup, however they are generally more accurate. [2, 82] The gold standards and other derivatives of both sensor classes are briefly presented below.

#### Wearable Sensors

For wearable sensors, footswitches [53, 79] and foot pressure insoles [82] constitute the gold standard. Both have the perk of detecting foot-floor contact directly and are therefore often used to validate data of other sensors [2, 53, 86, 90, 50, 82]. A disadvantage in using footswitches or foot pressure insoles is the limited number of detectable gait events. Other wearable sensors are inertial measurement units (IMUs) [86, 82] with accelerometers and gyroscopes in single [7, 57, 90, 2] or combined [53] usage. One advantage of these sensors is the greater granularity of the gait cycle, while disadvantages are an increased computational load and the dependency of correct placement on the respective body parts, as the recorded acceleration has to be double integrated to extract the considered values.

Electroneurography [82], electromyography [16] and oscillators [92, 95] can also be used for recording gait data. However, these sensors are less popular due to their high complexity in signal acquisition and post-processing.

All presented wearable sensors are also used in combination with each other [76, 82, 71]. A major disadvantage of all wearable sensors is the encumbrance of the tested subject. Therefore, these sensors are less suitable for subjects with difficulty in walking. [50] Additionally, performance of wearable sensors degrades noticeably at slower walking speeds, which is usually the pace of individuals with pathological gait. [47]

**Non-wearable Sensors**

The gold standards for non-wearable sensors are optical marker-based systems, also known as three-dimensional motion capture systems [70, 17], and force platforms [82]. In optical marker-based systems, reflective marker are attached to landmark body positions of a subject. Recorded by cameras, these marker are translated into three-dimensional spatial positions. [48] Generally, the recorded data of an optical marker-based system allows the recognition of four gait phases, but does typically not provide a higher granularity. To reach a higher granularity, additional variables like marker velocity or joint angles can be computed from marker positions using an analysis software. [82, 69] While events in normal gait can be detected automatically, events in pathological gait are more difficult to detect and might even be annotated manually in more severe gait disorders. This also motivates the search for algorithms automating the time-consuming annotation. [51]
Force platforms can provide a similar functionality as footswitches or foot pressure insoles, as they can detect foot-floor contact via vertical physical force. In addition, they also provide spatial information about the point of force transmission. Results dependent heavily on the tread of the foot. A clean hit refers to a step, where only one foot contacts a force platform without stepping over the edges of said platform. Clean force plate data is often missing as feet overlap from one platform to another or from the platform onto the floor, especially when subjects with pathological gait are being tested. In such cases, it is impossible to achieve reliable results without repeating the experiment several times, which, on the other hand, can encumber the tested subjects and distort results. [35, 51, 50]

### 2.1.4 Artificial Neural Networks

Artificial neural networks (ANNs) are massively parallel computing systems consisting of a large number of simple processing elements called neurons. These neurons are based on the concept of biological neurons and can be described as linear functions. Within an ANN, neurons are arranged in parallel structures called layers. In general, an ANN consists of an input layer, an output layer and one or more other layers with unidirectional connections between them. Such an arrangement is the most basic model of an ANN and is referred to as *single-layer* or *multi-layer perceptron*. These types of ANNs are memory-less. [44, 4, 46] In contrast to conventional machine learning approaches, ANNs are extremely flexible and perform well, even outside their trained domain. In addition, they are characterized by robustness, fault and failure tolerance, the ability to handle imprecise information and low energy consumption. [46]

**Recurrent Neural Networks**

Recurrent neural networks (RNNs) are an extension of conventional artificial neural networks to exhibit temporal behavior. RNNs contain a so-called *recurrent hidden state* which activation is dependent on the activation of previous time steps. Through directed connections between units of an individual layer, the network output is fed back into the network input neurons. Therefore, RNNs are able to handle sequential data and can recognize and learn temporal dependencies. [20, 49] However, RNNs are difficult to train on long-term dependencies. For an RNN to learn, an error signal must be backpropagated through time. The magnitude of the error signal depends exponentially on the magnitude of the weights of the RNN. Because a RNN overwrites its content at each time step, the back-propagated signal tends to blow up or vanish very quickly, thus making the RNN unable to learn further information. In general, a RNN fails to learn in presence of time lags greater than 5 - 10 discrete time steps. [20, 31]

#### Long Short-Term Memory Recurrent Neural Networks

Long short-term memory recurrent neural networks (LSTM RNNs) are an improvement over general recurrent neural networks. In order to address the problem of vanishing or exploding error signals, LSTM RNNs incorporate gate functions into their state dynamics. LSTM RNNs maintain a memory of their content, which is updated by partially forgetting existing memory and adding new memory content. The extent of updating memory is modulated by these gate functions. The gate functions allow a better control of the error signal and thus allowing to continue learning over more time steps. [20, 15, 49] In addition to their ability to capture longer time dependencies than conventional RNNs, they also achieve better results in general. LSTM RNNs can learn to bridge dependencies of up to 1000 discrete time steps. [31, 80]

## 2.2 Existing Methods

For presenting a new approach in the field of gait analysis, a validation against an established gold standard is not enough. The approach has to be compared against other existing methods in order to prove its performance and superiority. In this section, different existing approaches for detection of gait events and estimation of gait phases are presented. These methods were chosen because of their actuality, as they have been published in the past ten years, or because of their popularity, as they are an often cited and established standard. In both cases, the inclusion criterion was, that these methods use the same evaluation metrics as this thesis, presented in Section 3.2.4, to allow a comparison with the proposed algorithm. The published results of all presented approaches of this section are used to evaluate the performance of the proposed algorithm in Chapter 5.

### 2.2.1 Gait Event Detection

In the following, existing methods are presented, whose methodology focuses on the detection of gait events. For detection of gait events, the time difference between a detected event and ground truth has to be as small as possible. Furthermore, algorithms must be generalized for subjects with both physiological and pathological gait patterns, as these patterns differ greatly. To allow a comparison with the proposed algorithm, only results including the detection of IC and FC during overground or treadmill walking are considered. Unless stated otherwise, presented methods are used for offline event detection in optical marker-based systems and have been validated via force platforms.

Desailly et al. (2009) [23] propose a so-called *high pass algorithm* (HPA), which uses kinematics of heel and metatarsal marker to detect gait events. For evaluation of this method, a data set consisting of 20 children with cerebral palsy and eight healthy adults was used. All gait data was recorded with a sample rate of 50Hz. For the data set with children, the absolute time difference was $17 \pm 15\,ms$ for IC and $18 \pm 17\,ms$ for FC, while mean time difference was $1 \pm 23\,ms$ for IC and $-2 \pm 25\,ms$ for FC. For the adult data set, the absolute time difference was $29 \pm 17\,ms$ for IC and $17 \pm 7\,ms$ for FC, while mean time differences were $27 \pm 19\,ms$ and $-14 \pm 12\,ms$ respectively. During trials, recorded gait speeds ranged from $0.67\,\frac{m}{s}$ to $1.23\,\frac{m}{s}$ for children and from $1.03\,\frac{m}{s}$ to $1.36\,\frac{m}{s}$ for adults.

Ghoussayni et al. (2004) [33] propose a method which detects gait events by setting a threshold on sagittal foot velocity. For evaluation, twelve healthy subjects walked at self-selected slow and normal speeds. The used optical marker-based system consisted of six cameras, all of which recorded data with 60Hz. According to Ghoussayni et al., 90% of all detected events were within 16.7 $ms$ compared to a visual inspection of video footage.

Hansen et al. (2002) [35] propose a method that detects gait events by computing sagittal displacement between the overall center of pressure and ankle marker. The method was evaluated with four healthy subjects at self-selected speeds and achieved a mean time difference of $7.5 \pm 5.83\,ms$ for IC and $18.08 \pm 9.83\,ms$ for FC. Eight cameras with a sample rate of 120Hz were used by Hansen et al.

Hreljac [41, 42] designed two approaches with colleagues. In the first approach (2000) [41], Hreljac developed together with Marshall the so-called *Hreljac-Marshall algorithm* (HMA). This method uses the sagittal displacement of heel marker and the coronal displacement of metatarsal marker to detect gait events. For two healthy subjects at self-selected speeds, the algorithm achieved an absolute time difference of 4.7 $ms$ for IC and 5.6 $ms$ for FC, as well as a mean time difference of 1.2 $ms$ for both IC and FC. The setup included four cameras with a sample rate of 60Hz. In the second approach (2000) [42], Hreljac designed together with Stergiou a method, that uses angular acceleration of foot and leg marker. For this approach, only a single camera with a 180Hz shsmple rate was used. In ten healthy subjects at self-selected speeds, the mean time difference was 2.4 $ms$ for IC and 2.8 $ms$ for FC. The root mean squared time difference was 4.5 $ms$ for IC and 6.9 $ms$ for FC.

Hsue et al. (2007) [43] propose two approaches for the detection of gait events in children with pathological gait. The used data set consisted of five hemiplegic and three diplegic children, all of which walked at comfortable speeds. Gait data was recorded with eight cameras at 60Hz. Both approaches concentrate on foot and heel marker. The first approach uses the anterior-posterior acceleration and achieved an absolute time difference of 9 $ms$ for IC and 20 $ms$ for FC. The second method uses the acceleration on the vertical axis. The absolute time difference was 15 $ms$ for IC and 25 $ms$ for FC.

Kidziński et al. (2019) [51] propose an approach with several LSTM RNNs to detect gait events in real-time. In their approach, a combination of three consecutive LSTM RNNs is used for each leg of the subject. A dataset of the *Gilette Children's specialty healthcare center*, consisting of 18153 trials with 9092 annotated events, was used for development and evaluation. Further information about the subjects in the data set is not given. All data was sampled with 120Hz. The walking speed in the evaluation set was $0.56\,\frac{m}{s}$ to $1.14\,\frac{m}{s}$. The input for the algorithm consists of the relative position between hip, knee, ankle and toe marker and the pelvis marker, as well as the position of the pelvis marker itself. In addition, velocity of hip, knee, ankle, toes and pelvis marker and the acceleration of the pelvis marker were used. The algorithm was able to recognize 99% of all IC and 95% of all FC in the evaluation set. The mean time difference for detected events was 18.3 $ms$ for IC and 12.5 $ms$ for FC.

Miller (2009) [67] proposes an approach in which he uses a feedforward neural network. The network uses sagittal plane position, velocity and acceleration of the heel and toe marker as well as the foot-floor angle, angular velocity and angular acceleration of said marker. For evaluation, a dataset of the *Mary Free Bed Rehabilitation Hospital* was used. The data consists of 50 patients, primarily with cerebral palsy, with walking speeds ranging from $0.82\,\frac{m}{s}$ to $1.22\,\frac{m}{s}$. Gait data of all subjects was recorded using a setup of ten cameras with a sample rate of 120Hz. The mean time difference was $-3.96 \pm 13\,ms$ for IC and $-6.05 \pm 14.5\,ms$ for FC.

O'Connor et al. (2007) [69] propose a so-called *foot velocity algorithm* (FVA). For detection of gait events, the algorithm uses the vertical velocity of the calculated foot center. The algorithm was evaluated with 54 healthy children and three children with spastic diplegia at self-selected speeds. Gait data was recorded via four cameras at 120Hz. For healthy children, the absolute time difference was $15\,ms$ for IC and $11\,ms$ for FC. The mean time difference was $16 \pm 15\,ms$ for IC and $9 \pm 15\,ms$ for FC. For children with pathological gait, the absolute time difference was $7 \pm 6\,ms$ for IC and $23 \pm 10\,ms$ for FC and mean time difference was $-3 \pm 9\,ms$ for IC and $-6 \pm 26\,ms$ for FC.

Osis et al. (2014) [70] propose a method which uses a threshold on angular sagittal acceleration of ankle, knee and hip marker. The setup for recording marker trajectories consisted of eight cameras with a sample rate of 200Hz. For validation, 154 annotated subjects with walking speeds ranging from $2.65\,\frac{m}{s} \pm 0.22\,\frac{m}{s}$ were used from a database. No direct information is given on the results, but according to Osis et al. $89\% - 94\%$ of all detected events were within $20\,ms$.

Zeni et al. (2008) [94] propose two methods for the detection of gait events. Both methods were validated with a data set consisting of seven healthy subjects, seven subjects with multiple sclerosis (MS) and five stroke survivors. Gait data of all subjects was recorded with six cameras at 60Hz. The first method is described as coordinate-based and uses the displacement on the vertical axis between heel and sacrum marker as well as toe and sacrum marker, while the second method, described as velocity-based, uses the time of change of the anterior-posterior velocity of heel and toe marker. For the coordinate-based method, the mean time difference was $-16.2\,ms$ for IC and $5.7\,ms$ for FC for the healthy set, $34.8\,ms$ and $11\,ms$ for the MS set and $6.1\,ms$ and $-24.8\,ms$ for the stroke set. For the velocity-based method, the mean time difference was $-5.2\,ms$ and $3\,ms$ for the healthy set, $-8.9\,ms$ and $5.2\,ms$ for the MS set and $9.4\,ms$ and $-1.7\,ms$ for the stroke set.

### 2.2.2 Gait Phase Estimation

In the following section, existing methods are presented whose methodology focuses on the estimation of gait phases. Like with the detection of gait events, it is aimed to achieve the lowest possible difference between estimated phase and ground truth. For a comparison with the proposed algorithm, only results including the estimation of percentual gait phases during overground or treadmill walking are considered. The presented existing methods use mainly wearable devices for recording gait data. Unless stated otherwise, presented methods are used for offline estimation of gait phases.

Jiang et al. (2018) [47] propose an approach using linear discriminant analysis on myogram data of forearms and lower extremities. Using this approach, four different gait phases can be detected. For validation, nine healthy volunteers walked at $0.27\,\frac{m}{s}$, $0.4\,\frac{m}{s}$ and $0.55\,\frac{m}{s}$ on a treadmill. In comparison with a high-speed camera, average temporal error of gait phase detection was $55.2\,ms$, which translates into 2.1%.

Mariani et al. (2013) [58] use two IMUs to detect three sub phases in the stance phase. The IMUs were fixed on the forefoot and recorded pitch angular velocity of shank and foot. For evaluation, ten healthy subjects and 32 subjects with different degrees of ankle osteoarthritis walked at self-selected speeds. In comparison with foot pressure insoles, an averaged difference of $0.6 \pm 1.5\%$ was achieved.

Senanayake et al. (2010) [76] propose a method, which uses fuzzy logic for real-time estimation of gait phases. As input for the fuzzy logic, data of four foot pressure insoles is used in combination with angles of knee joints, recorded by two IMUs placed on the knee. The approach was validated on two healthy subjects at self-selected speeds. It was able to detect seven gait phases. The overall averaged difference was $1.26 \pm 6.08\%$.

Trojaniello et al. (2014) [83] designed a method which uses two magnetic and inertial measurement units (MIMUs) attached at the subject's ankles. Using angular velocity in the sagittal plane, the method is able to identify two gait phases. For evaluation, ten healthy subjects, ten hemiparetic subjects, ten subjects with choreic movement disorder and ten subjects with Parkinson's disease walked on a force platform at self-selected comfortable and higher speeds. The averaged mean absolute error for stance and swing duration at both speeds was 0.04 and 0.07 for healthy subjects, 0.06 for both stance and swing time in hemiparetic subjects, 0.03 and 0.055 for subjects with choreic movement disorder and 0.035 and 0.075 for subjects with Parkinson's disease.

Vu et al. (2018) [86] propose a so-called *exponentially delayed fully connected neural network* (ED-FNN) to estimate gait phases using sagittal angular velocity and sagittal angular acceleration of the foot. The data is recorded by an IMU placed on the lower shank. It was tested on seven healthy subjects at advanced gait speeds of $2.2\,\frac{m}{s}$, $2.6\,\frac{m}{s}$, $3.2\,\frac{m}{s}$ and $3.8\,\frac{m}{s}$. Foot pressure insoles conducted as a reference. The mean squared error was $0.008 \pm 0.001$ and the mean absolute error was $0.032 \pm 0.002$.

Yan et al. (2015) [92] propose a method for estimating the gait phase in real-time through adaptive oscillators. This method was tested on six healthy subjects, who walked at $0.67\,\frac{m}{s}$ to $1.22\,\frac{m}{s}$ while wearing an active pelvis orthosis. Using the hip joint angle and the vertical ground reaction force as a reference, an average root mean squared error of 0.011 was achieved.

Zheng et al. (2017) [95] propose an approach for estimating gait phases in real-time. This approach uses capacitive sensing oscillators to measure muscle deformation on shank and thigh. The approach was tested on three healthy subjects on self-selected slow, normal and fast walking speeds. Foot pressure insoles were used as a reference. The root mean square of the estimation error within stride was 0.041.

# 3 Material and Methods

This chapter presents the data set used for development of the proposed algorithm and the processing steps of said data set prior development. Furthermore, the development process and the functionality of individual components of the algorithm are presented, as well as selected methods for evaluation and used evaluation metrics.

## 3.1 Data

This section provides an overview about the subjects of the investigated data set, the experimental setup and protocol, and the annotation process. Additionally, the used digital filter and the process of feature selection are presented.

### 3.1.1 Subjects

For developing the proposed algorithm, the data set of *RehaGait* [75] was used. This data set consists of 23 measurements. Each measurement corresponds to an individual participant whose gait data was recorded during the project. The participants were 13 subjects with incomplete spinal cord injury (iSCI) and ten stroke survivors. Subjects with incomplete SCI were further divided into ten subjects with low impairment (iSCI L) and three subjects with moderate impairment (iSCI M). This distinction was carried out by an expert of gait analysis. Furthermore, all subjects were identified as well walking subjects with robust walking function according to their corresponding diagnosis. For generalization, eleven non-disabled subjects were selected as an additional reference for this thesis. Their gait data was recorded prior *RehaGait*.
The clinical set and the non-disabled, healthy set form the data set used to develop this algorithm. In this thesis, the combined data set of non-disabled subjects and patients is referred to as *complete set* (CS). A distribution of the subject features is shown in Table 3.1.

Table 3.1: Distribution of subject features in the investigated data sets

| | healthy set | clinical set | | | complete set |
|---|---|---|---|---|---|
| | | iSCI L set | iSCI M set | stroke set | |
| **age [y]** | 28 ± 3.5 | 52.2 ± 21.8 | 63 ± 8.4 | 57.1 ± 8 | 38.1 ± 17.2 |
| **height [cm]** | 177.5 ± 10.1 | 168 ± 7.4 | 167.3 ± 9.5 | 173.2 ± 12.1 | 174.3 ± 10.4 |
| **weight [kg]** | 72.7 ± 13.1 | 68.2 ± 13.3 | 61 ± 4 | 82.1 ± 17.4 | 73 ± 14.2 |
| **sex [male:female]** | 6 : 5 | 3 : 7 | 0 : 3 | 6 : 4 | 15 : 19 |

### 3.1.2 Experimental Setup and Protocol

All gait data of the healthy and clinical set was recorded in a separate laboratory of the *Spinal Cord Injury Center of Heidelberg University Hospital*. The experimental setup included an optical marker-based motion analysis system consisting of eight *Hawk* cameras

[37], manufactured by *MotionAnalysis (Santa Rosa, CA 95403 USA)*, a treadmill and 24 reflective markers. The markers were placed on anatomical landmarks in accordance with the Helen Hayes marker set [48], as illustrated in Figure 3.1. The Helen Hayes marker set was used, as it represents a compromise between the number of markers and the possibility of calculating the angles of all major joints based on the Helen Hayes biomechanical model [48]. The treadmill was located in the center of the laboratory, while the cameras were mounted on the ceiling and placed along the horizontal and diagonal of the treadmill. According the experimental protocol, the gait of all participants was measured for 90 seconds while they walked on the treadmill. Subjects of the clinical set walked at self-selected speeds, ranging from $0.2\,\frac{m}{s}$ to $0.8\,\frac{m}{s}$. For subjects of the healthy set, three individual measurements were taken with $0.2\,\frac{m}{s}$, $0.6\,\frac{m}{s}$ and $1\,\frac{m}{s}$. A total of 56 measurements were recorded. Gait data was recorded with a sample rate of 60Hz. One subject of the iSCI L set had to be excluded from the complete set, as its only recorded measurement had a sample rate of 200Hz. For each measurement, a total of 5400 individual frames were acquired. Each frame consists of the three-dimensional trajectory (TRB) data of the attached markers, as well as the hierarchical translation-rotation (HTR) data of the native skeleton segments computed by the motion capture software *Cortex* of *MotionAnalysis (Santa Rosa, CA 95403 USA)* using the Helen Hayes biomechanical model.

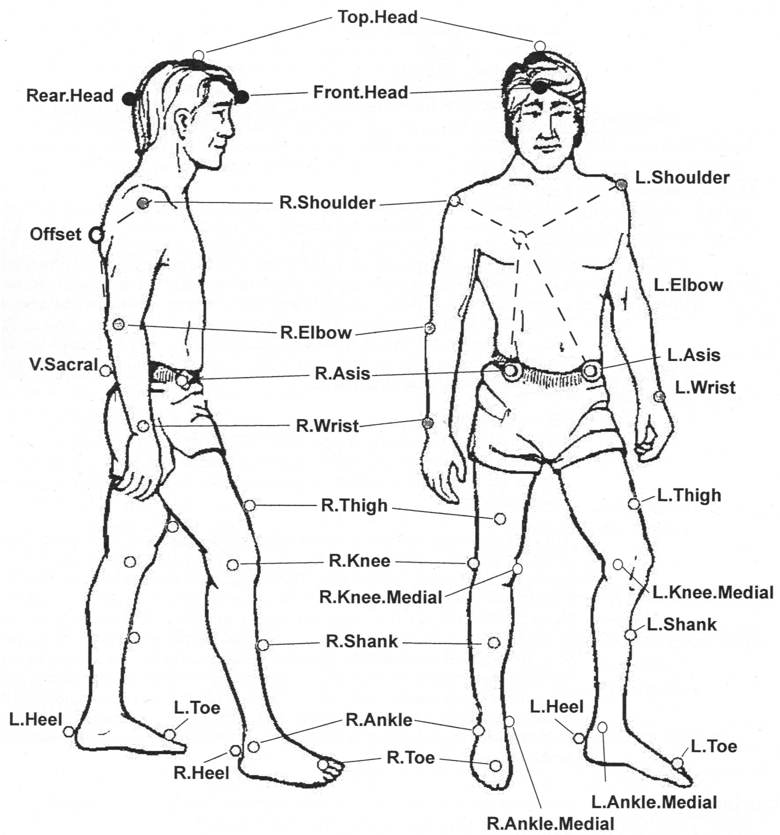


FIGURE 3.1: Helen Hayes marker set marker placement [68]

### 3.1.3 Data Annotation

An objective of this thesis is the evaluation of the self-developed algorithm used by the *Spinal Cord Injury Center of Heidelberg University Hospital*. Since no means of validation like footswitches or force plates were available, the recorded data was annotated manually by the author. For the annotation, the marker trajectories and computed segments of each measurement were visualized in *Cortex*. Using the visualized markers and segments, the occurrence of gait events was determined. As IC and FC are the most elementary gait events [10, 48, 17, 25] and the self-developed algorithm of the *Spinal Cord Injury Center* also concentrates on the detection of these, annotation was done only for IC and FC. The annotation was performed on the complete set.
As an annotation by a single rater is very subjective, further iterations of annotation were made. In these iterations, only a subset of the CS was annotated. The subset consisted of all measurements from the clinical set and 23 randomly selected measurements from the non-disabled, healthy set (69% of the healthy set). This quantity was chosen to balance the measurements between non-disabled subjects and patients. Due to time constraints, only the first 40 strides of each measurement were annotated in the subset. Since the annotation via *Cortex* was also time consuming, a *MATLAB* script for annotating gait events was created by a colleague of the *Spinal Cord Injury Center of Heidelberg University Hospital*. Screenshots of the respective annotation environments are shown in Figure 3.2 and 3.3.

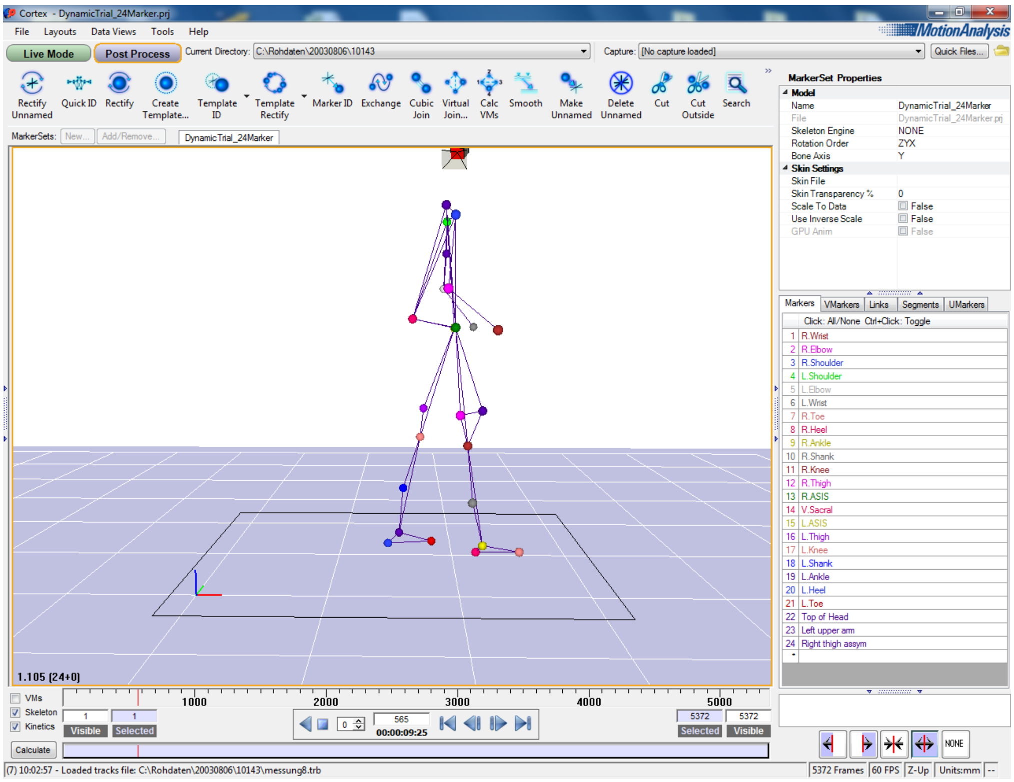


FIGURE 3.2: Screenshot of the gait analysis software *Cortex*. The main view visualizes all markers of the current frame. Frames are freely selectable in the bottom panel. The right panel shows further information of the visualized markers with the possibility to display past trajectories of selected markers in the main view

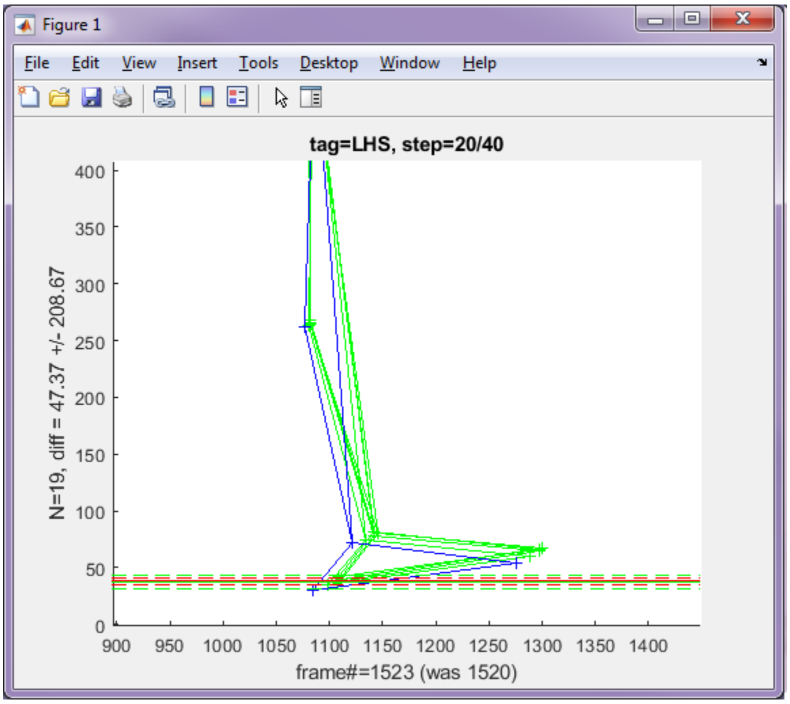


FIGURE 3.3: Screenshot of the self-developed script for manual annotation of gait events. Blue lines visualize the marker trajectories of the current frame. Green lines visualize marker trajectories of past frames. Horizontal lines represent the floor.

Using this script, the subset was annotated by the author and two colleagues of the *Spinal Cord Injury Center*, both of whom are experienced researchers working for several years in the field of gait analysis. In this thesis, these annotators are referred to as *Rater A*, *Rater B* and *Rater C*, whereas *Rater A* is the author of this thesis. Two iterations of annotation were carried out to allow a more profound investigation of the inter-rater and intra-rater reliability. Thus, the CS was annotated once by *Rater A*, while the subset was annotated twice by *Rater A*, *Rater B* and *Rater C*.

The inter-rater reliability is the extent of consistency among different observers [65, 77] and was determined by comparing the annotated subset of one rater with every other annotated subset of the same iteration. For intra-rater reliability, which is the extent of consistency among repeated observations by a single rater [14, 54], the annotated subsets by *Rater A* were compared with the annotated CS. Intra-rater reliability was only determined for *Rater A*, as *Rater B* and *Rater C* did not conduct an annotation of the CS.

For both inter-rater and intra-rater reliability, a Bland-Altman (BA) analysis [13, 12] via Bland-Altman plots was performed. In a BA plot, the mean values of two observations are plotted against the mean difference of the observations. From these plots the mean and the 95% limits of agreement (LOA), specified as $mean \pm 1.96 * standard\, deviation$, are obtained to quantify the agreement between two selected observations.

For a more profound BA analysis, the observations of the subsets were considered separately for IC and FC. Furthermore, within the BA analysis, the time difference between two consecutive annotated events, and not the times of the events themselves, were examined. In addition to measuring the agreement, this made it possible to examine the walking speed on the extent of agreement between observations. Measurements were marked according to their walking speed and subject's diagnosis with unique symbols and colors in the BA plots.

To further evaluate the degree of agreement, the correlation coefficients *Cohen's kappa*

[14], *Fleiss' kappa* [14] and *intraclass correlation coefficient* (ICC) [65, 77] were calculated. For intra-rater reliability, *Cohen's kappa* was calculated, while for inter-rater reliability, *Fleiss' kappa* and the two-way mixed, average score ICC were calculated. Different types of kappa were selected, as *Cohen's kappa* measures agreement between two raters only. For measuring agreement between more than two raters with kappa, *Fleiss' kappa* has to be used. Because kappa tends to overestimate agreement among raters, ICC was additionally computed for inter-rater agreement to further evaluate the scores [14]. The model of the ICC was selected using the guideline of Koo and Li [52]. For both kappa and ICC, p-values below 0.05 were considered significant. Kappa scores for agreement are interpreted as *slight* (<0.21), *fair* (0.21 - 0.4), *moderate* (0.41 - 0.60), *substantial* (0.61 - 0.80) and *perfect* (>0.81) [55], while ICC scores are interpreted as *poor* (<0.50), *moderate* (0.50 - 0.75), *good* (0.75 - 0.90) and *excellent* (>0.90) [52].
For calculation of the correlation coefficients, scripts by Cardillo G. [28, 29] and Arash S. [74] of the *MathWorks File Exchange* were used. Similar to the BA analysis, the observations were considered separately for IC and FC. In addition, a further separation of the observations was made based on the diagnosis of the subjects. The reason for this was to examine the extent of agreement in dependence of the subject's diagnosis, since a BA analysis needed further plots to get this insight, which was considered to be too time-consuming. As the real ground truth was unknown, a tolerance of $50\,ms$ was considered for the calculation of the correlation coefficients. Tolerance was chosen to be $50\,ms$ as deviations above $50\,ms$ were considered not tolerable and $50\,ms$ translate into exactly three frames with the given sample rate. To account for the tolerance, the average time difference and the root mean squared difference between the annotated events of two observations were calculated. After determining the degree of agreement, the median of all individual observations formed the ground truth for the development and evaluation of the proposed algorithm.

### 3.1.4 Filter Selection

When recording marker trajectories, high-frequency artifacts can appear. The origin of these artifacts can be movement of soft tissue or a pathological gait pattern [36, 30, 1]. To achieve better results, many approaches include digital filters to remove this noise. For optical marker-based systems, the Butterworth filter [19] is by far the most popular filter [17, 23, 25, 33, 38, 41, 66, 67]. Accordingly, a Butterworth filter is also used in this algorithm. To find a suitable configuration, different filter parameters were tested out in several iterations on all HTR and TRB data of two randomly selected measurements of the healthy and clinical set and assessed visually. The best result was a second order low-pass Butterworth filter with a cut-off frequency of 4Hz. The configuration was chosen, as it contains the best balance between smoothing of artifacts and preservation of relevant structures. Figure 3.4 displays a section of the applied filter on TRB and HTR data of a subject from the iSCI L set. The phase shift for the filtered data is five frames. As gait data is processed in real-time by the proposed algorithm, filtering must take place unidirectionally and therefore the phase shift has to be condoned. The shown filtering is equal to the filtering of all HTR and TRB data.

### 3.1.5 Feature Selection

For detecting gait events or regressing gait phases, each recorded marker or computed segments represents a feature dimension. With increasing dimensions, the corresponding algorithm is at risk of being subject to the so-called *curse of dimensionality* [8]. The curse of dimensionality is a term for an assortment of challanges presented in high-dimensional

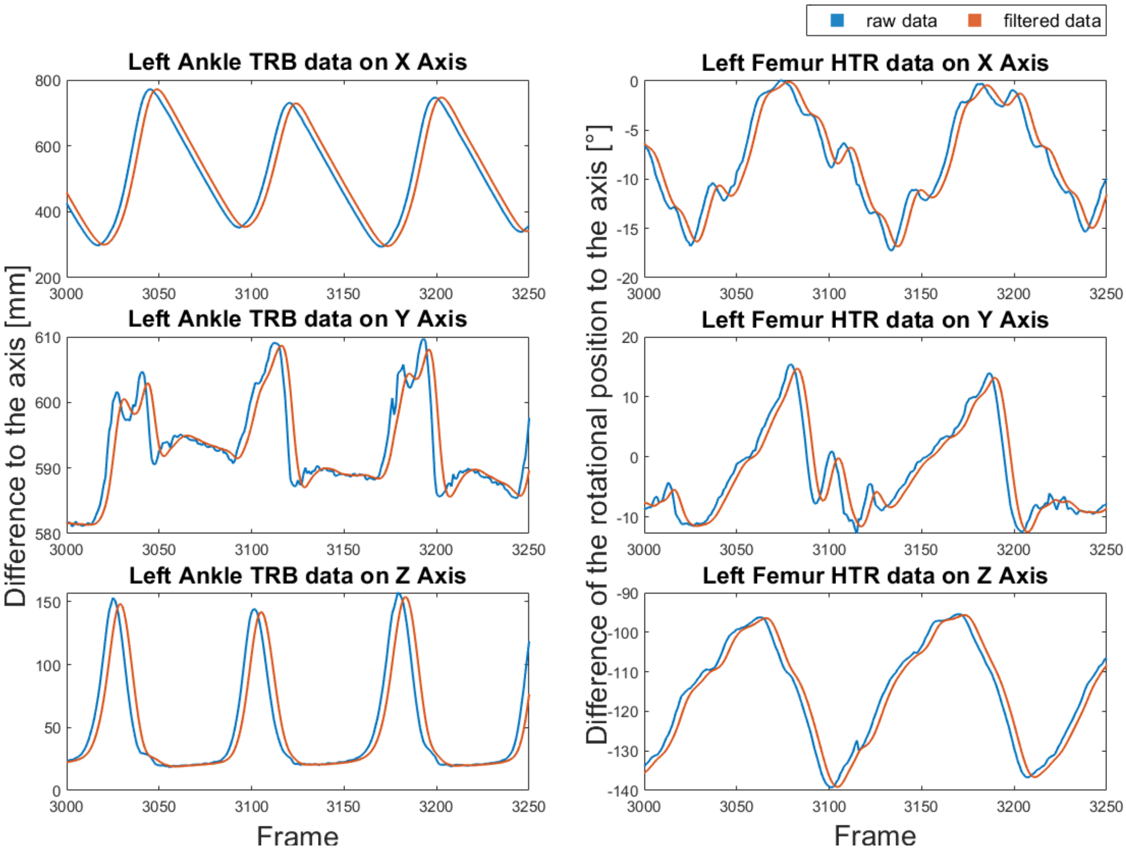


FIGURE 3.4: Exemplary filtering of gait data with a second-order Butterworth filter at a cut-off frequency of 4Hz. See Table 2.1 to translate axis into fundamental direction of body movement

spaces like incomparable scores, exponential search spaces or a significant high number of irrelevant attributes [96]. It is beneficial to reduce the number of dimensions, thus the number of considered features. With avoidance of the curse of dimensionality, the algorithm becomes simpler to handle and less computational load arises while processing data. In order to distinguish important features from unimportant ones, expert opinions may be used or established procedures from the field of feature selection [22].

In this thesis, a random forest [39] was used for the selection of features. A random forest is a collective of individual decision trees and is used for classification or regression tasks. Although this is not a feature selection technique, a random forest can be used for the selection of features, as more important features are put in the upper branches of a decision tree during training, while less important features are located in the lower branches. Getting insight on the location of the different features after training a random forest, one can infer the importance of said features.

The trajectories of recorded markers and the segments computed by *Cortex* made up a total of 132 individual features that changed over time. Each feature describes a position or an angle on an axis of a marker or segment. In order to reveal all possible information in the gait data, velocity and acceleration of each feature were calculated for each frame. Additionally, if two features were located on the same axis, the distance between those features on said axis was calculated for each frame. This information was extracted, as existing methods [84, 70, 71, 56, 38] also use this kind of information. Overall, the number of features increased from 132 to 2174.

The feature selection was done in *MATLAB Version 9.6*. For each case of application, detection of gait events and regression of gait phases, and each body side, a random forest with 100 decision trees was trained, thus a total of four times. Using the function *templateTree* [64], a tree template was created and trained for detection of gait events with the function *fitcensemble* [60]. Training a tree template for regression of gait phases was done using the function *fitrensemble* [61]. The default settings remained unchanged in each function. To train a random forest for detection of gait events, the ground truth from the annotation process was used. To create a ground truth for the regression of gait phases, the gait phases of each body side were computed using the time intervals between annotated ICs. The feature set was divided prior training into a development set and an evaluation set. From the whole feature set, 85% of all frames were added to the development set, while the remaining 15% were added to the evaluation set. The allocation was done randomly for each frame and enabled an evaluation of selected features after training. Due to time restraints, feature selection was conducted only once and only with the presented data set.

TABLE 3.2: Selected features for detection of gait events on one body side. See Table 2.1 to translate axis into fundamental direction of body movement

| **Body side** | **Marker/Segment** | **Feature** | **Axis** | **Feature Importance** |
|---|---|---|---|---|
| ipsilateral | heel | velocity | Z | 2.81 |
| | | acceleration | Z | 3.09 |
| | toe | velocity | XZ | 2.03 |
| | | acceleration | XZ | 1.94 |
| | ankle | velocity | X | 1.63 |
| | | acceleration | X | 2.74 |

TABLE 3.3: Selected features for regression of gait phases on one body side. See Table 2.1 to translate axis into fundamental direction of body movement

| **Body side** | **Marker/Segment** | **Feature** | **Axis** | **Feature Importance** |
|---|---|---|---|---|
| ipsilateral | heel | position | Z | 1.86 |
| | | velocity | Z | 1.09 |
| | | difference to contralateral heel | Z | 1.51 |
| | | difference to contralateral toe | Z | 1.34 |
| | toe | position | Z | 0.86 |
| | | velocity | Z | 4.34 |
| | ankle | position | Z | 1.59 |
| | | velocity | XZ | 1.32 |
| | | difference to contralateral toe | X | 0.69 |
| | | difference to contralateral ankle | Z | 1.22 |
| | shank | velocity | Z | 0.91 |
| contralateral | heel | velocity | X | 0.84 |
| | patella | position | Y | 0.98 |
| | tibia | difference to ipsilateral patella | Y | 0.87 |

For training a classification model, the development set was additionally stratified. Reason for this was the circumstance, that over 95% of the frames in the development set contained the class *no event*. This class imbalance made the development set highly skewed, which led to the random forest being incapable of learning. To remedy this situation, the classes were re-balanced. Frames with the class *no event* were selected and discarded randomly until an equal distribution of frames with and without event arose.

For this purpose, 98% of frames with the class *no event* had to be discarded.
After all random forest models had been trained, the feature importance of classification models was calculated using the function *oobPermutedPredictorImportance* [62]. In contrast to the previously described procedure for determining feature importance, this function estimates predictor importance via permutation of out-of-bag predictor observations. For regression models, the feature importance was calculated with the function *predictorImportance* [63]. This function estimates the predictor importance by summing the mean squared error on every branch split and dividing the sum by the number of branch nodes. The average feature importance for the features of all random forest models was $0.255 \pm 0.212$. Features with the highest importance were selected and used for the training of narrower random forest models. It should be noted that, according to the calculated feature importance, features containing joint angles seemed to be less important than features containing position or distance data. The performance of the newly trained models was examined using the evaluation set. As evaluation metrics, the respective metrics from Section 3.2.4 were used. Iteratively, features with the highest importance were removed from or added to the training of the random forest models until the smallest possible feature set with equal or better performance than the initially trained models with all features was achieved for each case of application. It is important to note that the chosen feature sets for the same case of application were, adapted to the respective body side, almost identical. No features with joint angles were found in these sets.
In order to further reduce the number of features, an intersection of features of both body sides was chosen as the final feature set. Another evaluation showed that the performance remained the same. The final feature sets for each case of application are shown in Table 3.2 and Table 3.3.

## 3.2 Algorithm Framework

This section explains the structure and the individual components of the proposed algorithm. The development process and the selected methods for evaluation and used evaluation metrics are also presented.

### 3.2.1 Development

For development, the ground truth containing only the selected features was used. The metrics from Section 3.2.4 were used for evaluation of trained networks. All gait data was filtered unidirectional with the presented filter. Like in feature selection, data sets for the training of gait event detection were stratified. Additionally, to simulate temporal dependencies, the individual frames from all data set were converted into sequences. Each sequence contained the values of the considered frame and a fixed amount of past frames. Several new data sets were created, each with a different sequence length.
The development of the algorithm took place in *MATLAB Version 9.6* using the *Deep Learning Toolbox* [59]. First, a simple LSTM RNN model was trained with the standard configuration in *MATLAB*. In extensive iterative processes, various parameters and sequence lengths were gradually tested out to find the best network architecture and training parameters for the corresponding case of application. Also, for detection of events, feature sets of both body sides were combined to explore, whether the detection of events can be covered with one single LSTM RNN cell. After finding a suitable configuration, the same configuration was used for training a network of the opposing body side. A loss of performance was not detected. Configuration for event detection could not be adopted to networks for regression of gait phases due to poor performance.

Therefore, a suitable configuration for architecture and training had to be determined again in extensive iterative processes.
Since in the training of neural networks the weights are initialized randomly, multiple models were built of each LSTM RNN to find the optimal performance. The models with the best performance were chosen as the models for the algorithm. The used training configuration of the networks and the individual architectures are shown in Appendix A. Following insights could be acquired during development:

- Performance of event detection considering one body side is better than performance of event detection considering both body sides
- The performance for event detection is best when the sequence to be classified consists of 12 frames before the actual event and 2 frames after the actual event. Thus, the LSTM RNN classifies 2 frames retroactively. The performance for regression of gait phases is best, when the sequence has a length of 28 frames
- Performance for event detection and phase regression can be further increased if each frame is classified separately, and not as a whole sequence. In this case, the LSTM RNN model has to be updated after each classified frame
- When detecting gait events frame-wise, not individual events but entire segments of consecutive events are detected. A post-processing of the output is necessary to reduce the high amount of false positive events

### 3.2.2 Architecture

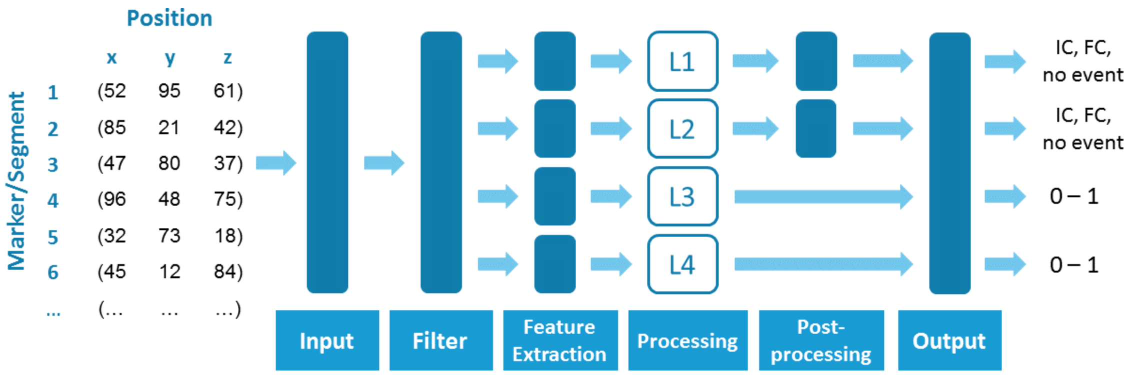


FIGURE 3.5: Pipeline of data processing within the proposed algorithm

Figure 3.5 shows the pipeline of data processing within the algorithm. The processing is as follows:

1. At each frame, the three-dimensional positions of both body sides for heel, toe, ankle and shank marker, as well as patella and tibia segment are read into the algorithm. In total, 36 values of 12 bodies are being tracked
2. All input data is filtered according to the presented filter
3. After filtering, the corresponding features are extracted from each body. To extract velocity and acceleration values, the last three corresponding features values are stored in an array in order to calculate first and second order differences. The extracted features are then grouped according to the required input of the respective LSTM RNN cell

4. The grouped features are processed within the individual LSTM RNNs. In Figure 3.5, the structures *L1* and *L2* are LSTM RNNs for the detection of gait events, while structures *L3* and *L4* are LSTM RNNs for the regression of gait phases. Regardless of the case of application, the processing time in each LSTM RNN cell is $42.29 \pm 2.76\,ms$. This processing time was achieved with an Intel i5-4760K CPU (3.4GHz) and 16GB RAM

5. Output of *L1* and *L2* is further processed. First, *L1* and *L2* are switched off for an fixed amount of five frames after the detection of an event. This is done to prevent the presented segmented detection of consecutive events. The amount was chosen, as no false positive event of the mentioned consecutive detection was found beyond this time instance. Distances between the last five detected events are stored in an array. If a new event is detected, the algorithm inspects, if the distance to the last detected events lies within a computed threshold. The threshold is calculated each time a new event is detected using the distance values within the array. If the distance between last detected event and newly detected event lies outside this threshold, the newly detected event is discarded. Otherwise, the event is considered viable and the oldest computed distance in the array is overwritten. This post-processing prevents the consecutive detection of false positive events after the detection of an event. It is based on the assumption, that the distances between events are within a certain scope when walking speed is constant. In addition, by overwriting the distances in the array, the algorithm can gradually adapt to a changing walking speed

6. After postprocessing, the results of *L1*, *L2*, *L3* and *L4* are passed as the final output of the algorithm for the current observed frame and can be used by a connected application for further processing. The final output consists of an assessment for both body sides, which kind of event the considered frame contains (*Initial Contact*, *Final Contact*, *no event*), and the current percental gait phase for both body sides, described as a number between 0 and 1

### 3.2.3 Evaluation Procedure

For the evaluation of the proposed algorithm, the self-developed and frequently used algorithm of the *Spinal Cord Injury Center*, abbreviated in the following parts of this thesis as *SCHOB*, and the approaches of Zeni et al. [94] were selected. The methods by Zeni et al. [94] were selected because they are one of the most popular methods to evaluate novel approaches [17, 25, 38, 56] in the field of gait event detection and are very easy to implement and use. Since two approaches were presented by Zeni et al. [94], a distinction was made between the coordinate-based approach (ZENI C) and the velocity-based approach (ZENI V). SCHOB uses heuristics for detection of gait events. As all data is bidirectional filtered prior processing, SCHOB is only suitable for offline detection of events. The complete set with the median of all annotated events was used as a ground truth. The median was chosen to reduce influence of events beyond the calculated limits of agreements of the BA plots.
For the evaluation of event detection, the proposed algorithm is abbreviated as *LSTM*. A tolerance of $50\,ms$ was considered for detected events of all methods. For the evaluation of phase estimation, the LSTM RNNs for regression of gait phases is abbreviated as *LSTM P*. As an objective of this thesis is to investigate, which case of application is better suited for determining gait phases, the LSTM RNNs for detection of gait events (LSTM E) were also considered in this comparison. As in the course of the research no approach was found for real-time gait phase estimation in optical marker-based systems,

no method was implemented for this. Instead, the methods of Zeni et al. [94] were used. All gait phases were calculated from time intervals of detected events, including the ground truth. Since the phases were calculated after the detection of all gait events, gait phases were not estimated under real-time conditions. The exception is LSTM P, as corresponding gait phases were computed directly from the input.

### 3.2.4 Evaluation Metrics

The performance of the proposed algorithm was evaluated using common machine learning metrics for classification [40] and regression [91, 18]. For detection of gait events the metrics *sensitivity* and *precision* were used. Since most of recorded data contains no gait events, more focus was put on finding and correctly classifying frames with events. Timing differences between detected events and ground truth were evaluated using the *mean time difference* and *root mean squared time difference*. To evaluate the estimation of gait phases, the metrics *mean squared error* and *mean absolute error* were used. Table 3.4 describes the presented metrics with their corresponding formula. In Table 3.4, TP, FP, FN, $t_e$, $\hat{t}_e$, $p_i$ and $\hat{p}_i$ correspond to:

- True positives (TP) : The number of positive test results correctly classified as positive results regarding the classification of a single class
- False positives (FP) : The number of negative test results incorrectly classified as positive results regarding the classification of a single class
- False negatives (FN) : The number of positive test results incorrectly classified as negative results regarding the classification of a single class
- Determined time of event ($t_e$) : Determined event time for the event $e$
- Actual time of event ($\hat{t}_e$) : Event time in ground truth for the event $e$
- Determined phase ($p_f$) : Determined percentual gait phase for the frame $f$
- Actual phase ($\hat{p}_f$) : Percentual gait phase in ground truth for the frame $f$

Table 3.4: Performance metrics used for evaluation of the proposed algorithm

| **Metric** | **Expression** | **Description** |
|---|---|---|
| sensitivity | $\frac{TP}{TP+FN}$ | proportion of correctly classified frames regarding the classification of a single class |
| precision | $\frac{TP}{TP+FP}$ | proportion of actual frames with class in classified frames regarding the classification of a single class |
| mean time difference | $\frac{1}{n}\sum_{e=1}^{n}\left(t_e - \hat{t}_e\right)$ | averaged difference between determined times of events and actual event times |
| root mean squared time difference | $\sqrt{\frac{1}{n}\sum_{e=1}^{n}\left(t_e - \hat{t}_e\right)^2}$ | square root of the averaged squared difference between determined times of events and actual event times |
| mean squared error | $\frac{1}{m}\sum_{i=1}^{m}(p_f - \hat{p}_f)^2$ | averaged squared difference between estimated phases and actual phases |
| mean absolute error | $\frac{1}{m}\sum_{i=1}^{m}\lvert p_f - \hat{p}_f\rvert$ | averaged absolute difference between estimated phases and actual phases |

# 4 Results

This chapter presents the evaluation results of the proposed algorithm regarding detection of gait events and estimation of gait phases. Results contain evaluation metrics and timing differences of selected methods to the ground truth. For manual annotation, the degree of inter-rater and intra-rater reliability is determined via Bland-Altman plots, correlation coefficients and temporal differences between observations.

## 4.1 Manual Annotation of Gait Data

This section presents the results of the manual annotation. The Bland-Altman (BA) plots are shown in Figure 4.1 and Figure 4.2. The plots display the time differences between annotated events of two different observations. Calculated values of mean, standard deviation and limits of agreement (LOA) for each BA plot are visible in Table 4.1. The LOA are calculated within a BA plot as $mean \pm 1.96 * standard\,deviation$ [13]. The results of the correlation coefficients *Cohen's kappa* [14], *Fleiss' kappa* [14] and *intraclass correlation coefficient* (ICC) [65, 77] are shown in Table 4.2. The mean and the root mean squared time differences between annotated events are shown in Table 4.3 and Table 4.4.

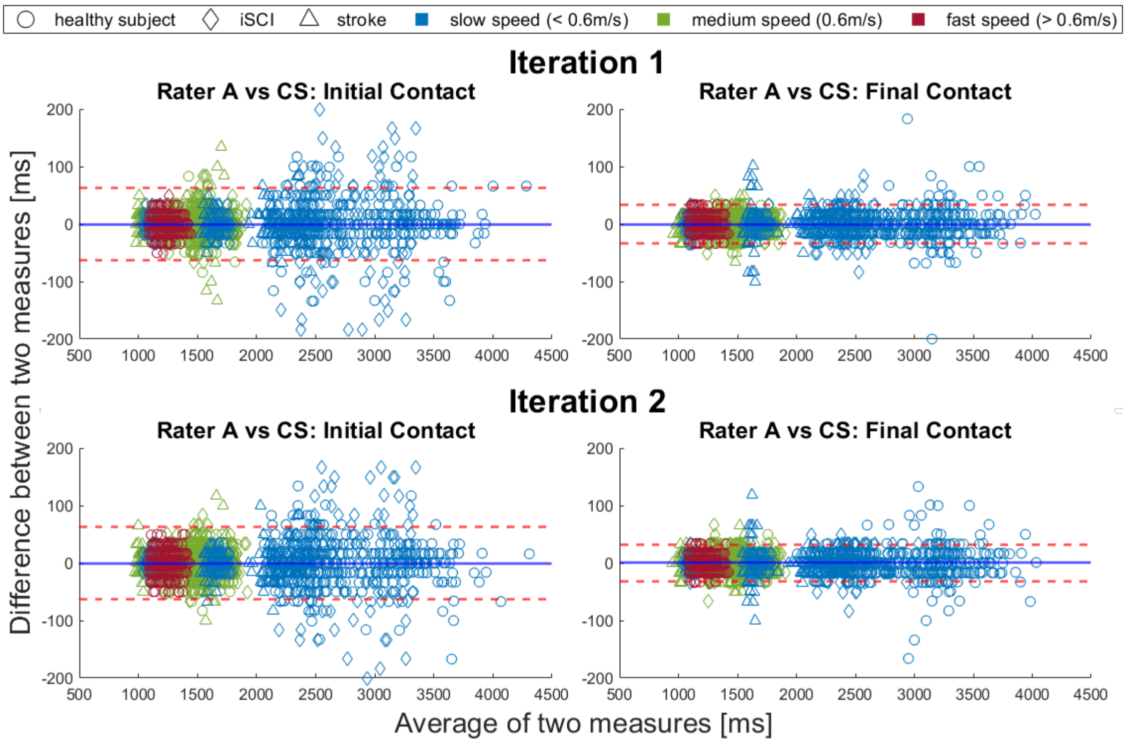


FIGURE 4.1: Bland-Altman plots for intra-rater reliability between Rater A and the complete set

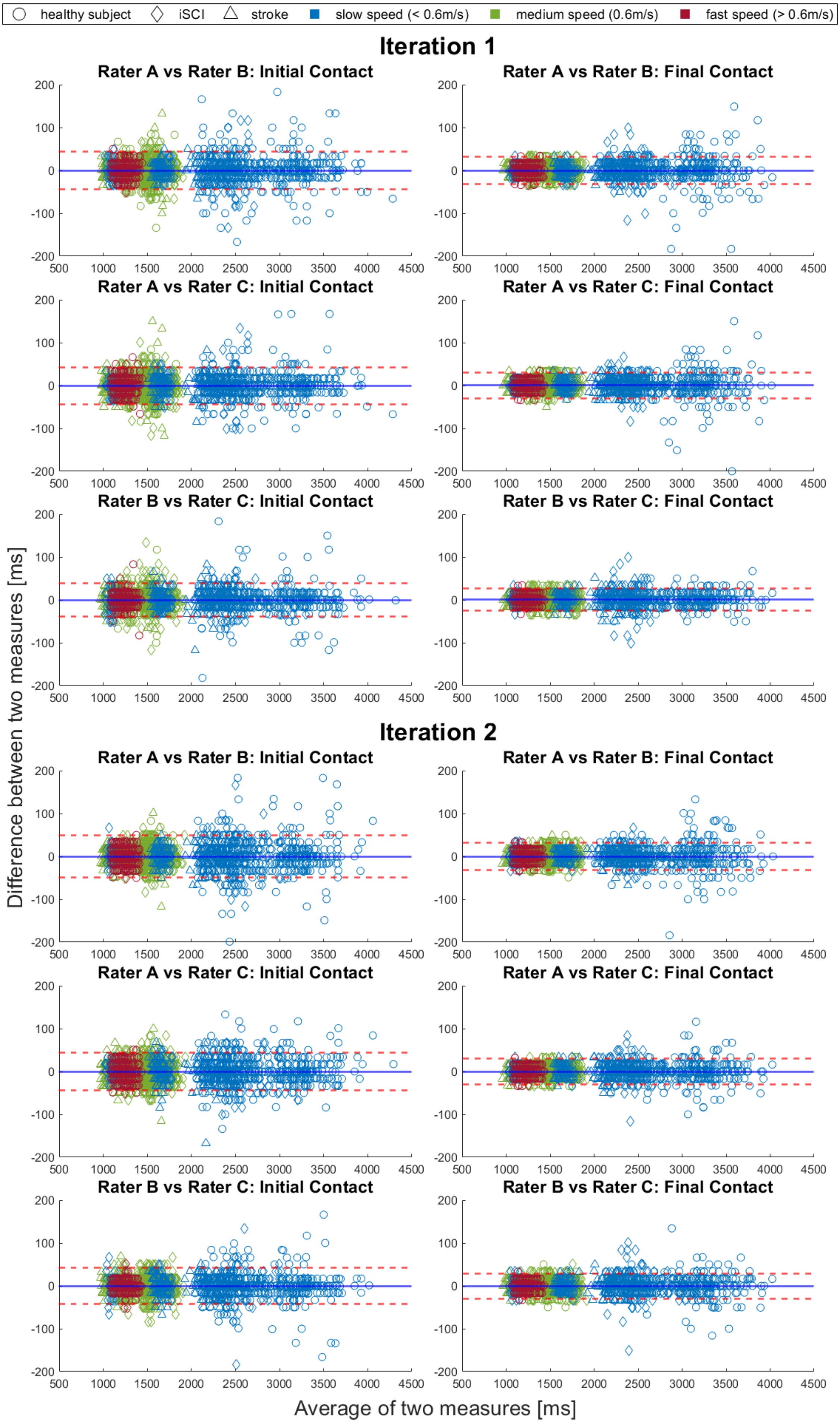


FIGURE 4.2: Bland-Altman plots for inter-rater reliability between Rater A, Rater B and Rater C

TABLE 4.1: Results of Bland-Altman analysis for intra-rater reliability between Rater A and the complete set, and inter-rater reliability between Rater A, Rater B and Rater C. Unit of all values is milliseconds [$ms$]

| Rater | Iteration | IC | | | | FC | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Difference | | Limits of Agreement | | Difference | | Limits of Agreement | |
| | | mean | std | lower | upper | mean | std | lower | upper |
| A vs *CS* | 1 | -0.203 | 32.078 | -63.077 | 62.670 | -0.085 | 17.198 | -33.794 | 33.624 |
| | 2 | -0.198 | 32.277 | -63.462 | 63.065 | 0.024 | 16.281 | -31.898 | 31.945 |
| A vs B | 1 | -0.029 | 22.815 | -44.747 | 44.688 | -0.014 | 16.050 | -31.472 | 31.444 |
| | 2 | -0.038 | 25.218 | -49.465 | 49.389 | -0.028 | 16.539 | -32.446 | 32.388 |
| A vs C | 1 | -0.136 | 21.985 | -43.227 | 42.955 | 0.064 | 15.049 | -29.431 | 29.560 |
| | 2 | -0.118 | 22.524 | -44.266 | 44.029 | -0.111 | 15.135 | -29.776 | 29.554 |
| B vs C | 1 | -0.106 | 19.491 | -38.310 | 38.096 | 0.078 | 13.231 | -25.855 | 26.013 |
| | 2 | -0.080 | 21.854 | -42.915 | 42.754 | -0.082 | 14.995 | -29.473 | 29.309 |

For all observations, the mean difference is less than 1 $ms$. The maximal standard deviation is 33 $ms$ for IC and 17 $ms$ for FC. In both intra-rater and inter-rater agreement, the LOA for FC are narrower than for IC. It is recommended by Bland and Altman, that 95% of the data points should lie within the LOA for an accordance of observations [13, 12]. The average percentage of outlying events for all BA plots is $4.31 \pm 0.63\%$. Therefore, this recommendation is fulfilled. Accordance between measurements is lower, if the respective participant has a slow walking speed. Regardless of the participant's diagnosis, measurements with slow walking speeds have the largest spread and their data points exceed the LOA more often, while measurements with medium and fast walking speeds have a comparable lower spread. Also, measurements with fast walking speeds are almost uniformly within the LOA. Except for the presented correlation of rater agreement and walking speed, no trend in data points can be detected.

TABLE 4.2: Results of descriptive statistics for intra-rater reliability between Rater A and the complete set, and inter-rater reliability between Rater A, Rater B and Rater C. Interpretations for correlations coefficients are located below the actual values. Interpretations *moderate*, *substantial* and *excellent* are abbreviated as *mod.*, *sub.* and *excell.* All p-values are below 0.05

| Descriptive Statistics | Iteration | healthy set | | iSCI L set | | iSCI M set | | stroke set | |
|---|---|---|---|---|---|---|---|---|---|
| | | IC | FC | IC | FC | IC | FC | IC | FC |
| Cohen's kappa | 1 | 0.948 | 0.995 | 0.944 | 0.999 | 0.457 | 0.929 | 0.937 | 0.991 |
| | | perfect | perfect | perfect | perfect | mod. | perfect | perfect | perfect |
| | 2 | 0.799 | 0.983 | 0.964 | 0.999 | 0.420 | 0.934 | 0.936 | 0.987 |
| | | sub. | perfect | perfect | perfect | mod. | perfect | perfect | perfect |
| Fleiss' kappa | 1 | 0.888 | 0.976 | 0.763 | 0.952 | 0.755 | 0.786 | 0.858 | 0.977 |
| | | perfect | perfect | sub. | perfect | sub. | sub. | perfect | perfect |
| | 2 | 0.581 | 0.940 | 0.744 | 0.942 | 0.877 | 0.792 | 0.812 | 0.983 |
| | | mod. | perfect | sub. | perfect | perfect | sub. | perfect | perfect |
| ICC | 1 | 0.959 | 0.992 | 0.906 | 0.984 | 0.895 | 0.912 | 0.947 | 0.992 |
| | | excell. | excell. | excell. | excell. | good | excell. | excell. | excell. |
| | 2 | 0.806 | 0.979 | 0.897 | 0.980 | 0.953 | 0.913 | 0.928 | 0.994 |
| | | good | excell. | good | excell. | excell. | excell. | excell. | excell. |

Table 4.3: **Mean time difference** of annotated events between Rater A and the complete set, and between Rater A, Rater B and Rater C. Standard deviation is located beneath each timing difference. Unit of all values is milliseconds [$ms$]

| Rater | Iteration | healthy set | | iSCI L set | | iSCI M set | | stroke set | |
|---|---|---|---|---|---|---|---|---|---|
| | | IC | FC | IC | FC | IC | FC | IC | FC |
| A vs CS | 1 | 12.28 | -5.81 | 18.85 | -5.85 | 243.55 | -12.57 | 14.78 | -82.98 |
| | | ±16.02 | ±7.75 | ±12.07 | ±6.39 | ±250.15 | ±19.14 | ±11.03 | ±241.07 |
| | 2 | 40.63 | -15.00 | 17.76 | -5.77 | 242.14 | -9.72 | 15.82 | -83.08 |
| | | ±14.41 | ±12.18 | ±10.54 | ±6.43 | ±241.68 | ±25.29 | ±13.13 | ±241.88 |
| A vs B | 1 | -19.41 | 8.06 | -29.77 | 26.52 | -38.51 | 46.01 | -22.00 | 16.06 |
| | | ±19.10 | ±28.14 | ±14.21 | ±15.39 | ±24.78 | ±15.30 | ±11.53 | ±10.29 |
| | 2 | -48.13 | 28.58 | -28.17 | 25.42 | -21.64 | 37.37 | -23.54 | 17.57 |
| | | ±17.71 | ±22.72 | ±9.89 | ±10.66 | ±25.65 | ±6.04 | ±10.41 | ±8.49 |
| A vs C | 1 | -32.15 | 5.29 | -47.91 | 1.76 | -27.72 | 8.59 | -41.77 | 0.02 |
| | | ±14.34 | ±22.49 | ±13.65 | ±5.08 | ±9.46 | ±3.45 | ±13.39 | ±6.31 |
| | 2 | -62.34 | 8.40 | -53.22 | -1.43 | -19.19 | -10.67 | -45.56 | -3.25 |
| | | ±15.59 | ±14.78 | ±11.51 | ±6.25 | ±13.00 | ±7.23 | ±13.34 | ±4.87 |
| B vs C | 1 | -12.74 | -2.76 | -18.14 | -24.77 | 10.78 | -37.42 | -19.77 | -16.04 |
| | | ±11.72 | ±10.88 | ±11.56 | ±14.78 | ±21.44 | ±17.53 | ±8.60 | ±11.26 |
| | 2 | -14.21 | -20.18 | -25.05 | -26.85 | 2.46 | -48.04 | -22.03 | -20.82 |
| | | ±9.78 | ±10.99 | ±3.49 | ±14.83 | ±12.69 | ±12.69 | ±8.94 | ±9.27 |

Table 4.4: **Root mean squared time difference** of annotated events between Rater A and the complete set, and between Rater A, Rater B and Rater C. Standard deviation is located beneath each timing difference. Unit of all values is milliseconds [$ms$]

| Rater | Iteration | healthy set | | iSCI L set | | iSCI M set | | stroke set | |
|---|---|---|---|---|---|---|---|---|---|
| | | IC | FC | IC | FC | IC | FC | IC | FC |
| A vs CS | 1 | 23.63 | 16.15 | 26.12 | 13.03 | 256.72 | 29.66 | 21.98 | 95.43 |
| | | ±16.15 | ±9.53 | ±10.75 | ±2.24 | ±251.67 | ±7.77 | ±10.93 | ±238.38 |
| | 2 | 46.89 | 21.91 | 24.45 | 13.21 | 254.64 | 30.35 | 24.01 | 96.80 |
| | | ±12.93 | ±10.77 | ±9.19 | ±2.29 | ±244.14 | ±10.80 | ±11.15 | ±237.99 |
| A vs B | 1 | 27.89 | 22.37 | 34.56 | 28.47 | 47.62 | 48.79 | 26.24 | 20.49 |
| | | ±20.58 | ±25.36 | ±12.96 | ±14.74 | ±24.90 | ±16.76 | ±12.02 | ±9.46 |
| | 2 | 53.76 | 31.79 | 32.42 | 27.43 | 40.35 | 40.00 | 28.64 | 21.69 |
| | | ±18.78 | ±23.86 | ±9.45 | ±10.16 | ±18.52 | ±6.77 | ±9.45 | ±7.33 |
| A vs C | 1 | 36.91 | 18.66 | 51.29 | 9.90 | 35.90 | 18.33 | 44.80 | 11.57 |
| | | ±14.74 | ±20.97 | ±12.88 | ±2.70 | ±8.24 | ±4.18 | ±14.15 | ±3.99 |
| | 2 | 65.85 | 15.66 | 55.78 | 10.12 | 32.45 | 21.95 | 49.06 | 12.22 |
| | | ±14.40 | ±14.99 | ±11.72 | ±2.98 | ±12.39 | ±7.47 | ±13.14 | ±3.32 |
| B vs C | 1 | 21.56 | 13.92 | 26.14 | 26.48 | 31.88 | 40.78 | 24.63 | 20.50 |
| | | ±8.29 | ±6.24 | ±4.87 | ±14.55 | ±9.98 | ±19.46 | ±8.21 | ±8.96 |
| | 2 | 21.75 | 23.61 | 29.19 | 28.96 | 23.71 | 51.88 | 26.81 | 23.85 |
| | | ±10.38 | ±11.54 | ±4.22 | ±13.87 | ±9.31 | ±13.35 | ±9.55 | ±7.70 |

The correlation coefficients show in general very high scores for agreement. For almost all observations, *perfect* or *excellent* agreement is achieved. Agreement for FC is consistently higher than for IC in all observations. It is also noticeable, that for *Fleiss' kappa*, only *substantial* agreement is achieved for IC in iSCI L and FC in iSCI M. Uniformly higher agreement is provided by ICC. Likewise, agreement in *Cohen's kappa* for IC in iSCI M is

only *moderate*. Agreement also differs considerably between iterations. All correlations coefficients for IC in the healthy set drop considerably in the second iteration. On the other hand, inter-rater scores for IC in iSCI M increase.
In temporal differences, the deviation between all observations is on average between 20 and 40 $ms$, which corresponds to about two frames. Especially noticeable is the IC of the iSCI M set and the FC of the stroke set in inter-rater agreement, where timing differences deviate by factor five in comparison with other observations. As in the correlation coefficients, the agreement for IC in the healthy set deteriorates in the second iteration. The observed improvement in inter-rater agreement for IC in the iSCI M set is also present.

## 4.2 Gait Event Detection

This section compares the performance of the proposed algorithm regarding the detection of gait events with other selected methods. For comparison of performance, the sensitivity and precision as well as the mean time difference and the root mean squared time difference of detected events to the ground truth were calculated. Results for event detection are shown in Table 4.5 and Table 4.6. Timing differences of detected events are displayed in Table 4.7 and Table 4.8.

Table 4.5: Comparison of **sensitivity** between selected methods and ground truth. Standard deviation is located beneath each sensitivity value. Unit of all values is percentage [%]

| **Algorithm** | | **healthy set** | | **iSCI L set** | | **iSCI M set** | | **stroke set** | |
|---|---|---|---|---|---|---|---|---|---|
| | | **IC** | **FC** | **IC** | **FC** | **IC** | **FC** | **IC** | **FC** |
| LSTM | | 93.77 | 99.22 | 99.30 | 99.43 | 85.85 | 97.70 | 97.01 | 99.77 |
| | | ±11.51 | ±2.48 | ±1.47 | ±1.12 | ±18.81 | ±3.26 | ±4.57 | ±0.48 |
| SCHOB | | 53.20 | 99.02 | 86.62 | 98.55 | 54.55 | 34.66 | 82.31 | 98.07 |
| | | ±34.11 | ±2.61 | ±15.08 | ±3.85 | ±36.36 | ±34.82 | ±24.36 | ±3.18 |
| ZENI | C | 90.00 | 96.81 | 93.12 | 99.91 | 72.88 | 81.22 | 73.86 | 98.33 |
| | | ±19.07 | ±6.16 | ±7.24 | ±0.25 | ±24.59 | ±22.01 | ±22.54 | ±3.97 |
| | V | 90.66 | 95.08 | 89.48 | 99.75 | 64.64 | 93.54 | 67.50 | 99.58 |
| | | ±18.66 | ±10.59 | ±10.58 | ±0.35 | ±37.32 | ±9.13 | ±24.91 | ±1.01 |

Table 4.6: Comparison of **precision** between selected methods and ground truth. Standard deviation is located beneath each precision value. Unit of all values is percentage [%]

| **Algorithm** | | **healthy set** | | **iSCI L set** | | **iSCI M set** | | **stroke set** | |
|---|---|---|---|---|---|---|---|---|---|
| | | **IC** | **FC** | **IC** | **FC** | **IC** | **FC** | **IC** | **FC** |
| LSTM | | 93.21 | 98.93 | 98.66 | 99.52 | 85.31 | 97.70 | 96.58 | 99.86 |
| | | ±12.85 | ±3.29 | ±1.94 | ±1.13 | ±18.46 | ±3.26 | ±4.11 | ±0.43 |
| SCHOB | | 52.78 | 96.47 | 86.52 | 98.55 | 54.48 | 34.66 | 82.31 | 98.16 |
| | | ±34.61 | ±12.93 | ±15.05 | ±3.85 | ±36.43 | ±34.82 | ±24.36 | ±3.22 |
| ZENI | C | 88.85 | 96.57 | 92.39 | 99.57 | 71.94 | 81.22 | 73.24 | 98.12 |
| | | ±19.61 | ±6.51 | ±7.12 | ±0.54 | ±23.90 | ±22.01 | ±22.43 | ±4.57 |
| | V | 88.82 | 95.02 | 88.50 | 100.00 | 63.31 | 93.54 | 66.65 | 99.67 |
| | | ±19.70 | ±10.83 | ±10.35 | | ±36.50 | ±9.13 | ±24.63 | ±1.00 |

An examination of the results for sensitivity and precision reveals, that LSTM provides almost the best performance for both healthy individuals and individuals with pathological gait. Detection rate is very high and dispersion is comparably very low. Only in the detection of FC in iSCI L, both methods of ZENI provide a better performance. However, the advantage is only slightly, whereas the performance of LSTM is on average 10 - 20% better than the other methods, most noticeable in the detection of IC in the iSCI M and the stroke set. For most events, LSTM has at least a performance of 95%, with a maximum of up to 99%. Performance of LSTM degrades only slightly in subjects of the clinical set. Both methods of ZENI have an overall slightly lower performance, with SCHOB having the lowest performance of all methods. Only in two cases is SCHOB superior to ZENI. Performance for detecting FC is on average higher than for detecting IC in all methods.

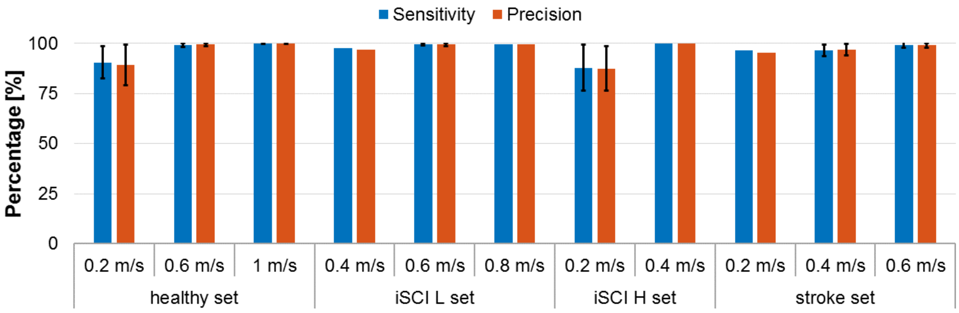


FIGURE 4.3: Averaged sensitivity and precision of proposed algorithm in relation to different walking speeds of investigated subjects. Bars without error indicator contain only a single measurement of corresponding combination of diagnosis and walking speed

Figure 4.3 shows the detection rate of LSTM, averaged for IC and FC, in relation to different walking speeds. Detection rate is very high and nearly consistent over the different pathologies and walking speeds. It is noticeable, that measurements with the walking speed 0.2 $\frac{m}{s}$ show lower detection rates and higher dispersion compared to other walking speeds of their corresponding data sets.

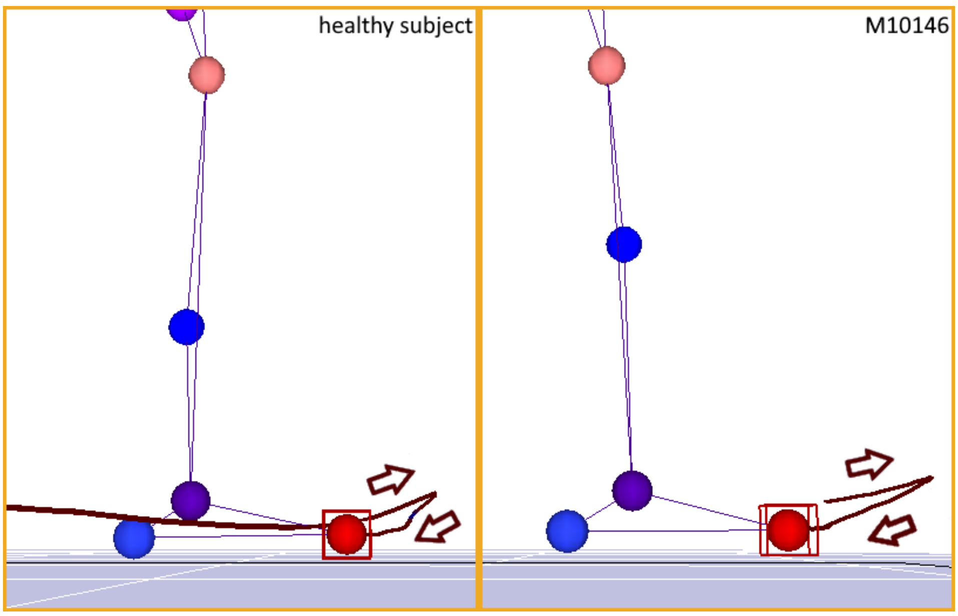


FIGURE 4.4: Comparison of left toe marker trajectories between healthy subject and measurement M10146. Both images show a time instance shortly after IC. The shown line and arrows indicate the course and direction of past marker trajectories

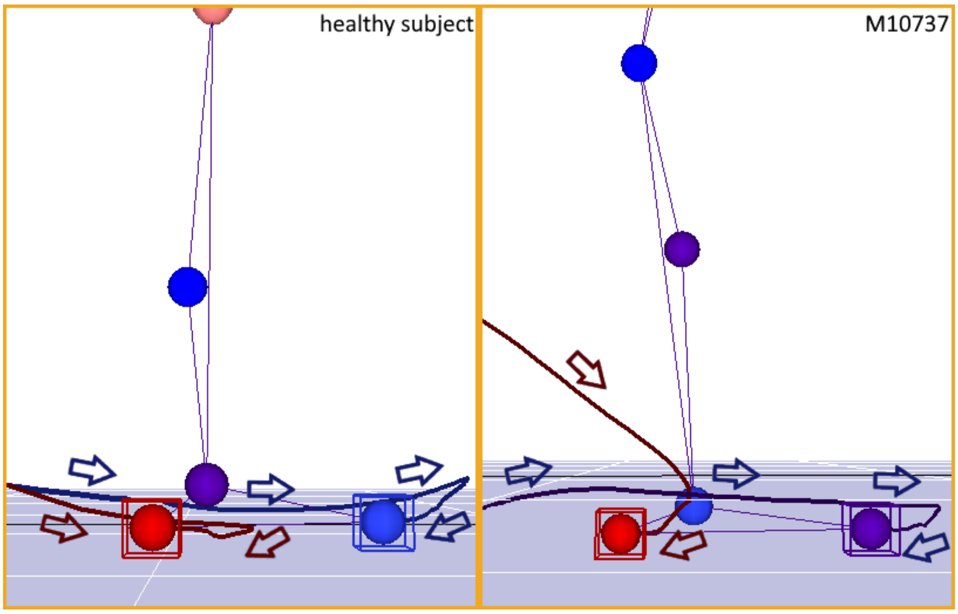


FIGURE 4.5: Comparison of left heel (red) and toe (blue) marker trajectories between healthy subject and measurement M10737. Both images show a time instance shortly after IC. The shown lines and arrows indicate the course and direction of past marker trajectories

It should be noted, that for two measurements, a high drop in performance could be observed in all methods. Only the detection of IC is affected. According to their subject identification number, these measurements are referred to as *M10146* and *M10737*. Figure 4.4 and Figure 4.5 show the course of gait trajectories during IC of these measurements in comparison with a healthy subject without drop in performance. Measurement M10146 contains gait data from a healthy subject with a walking speed of 0.2 $\frac{m}{s}$. Figure 4.4 reveals that, in comparison with a healthy subject, the foot in M10146 is further swung forward before initializing IC. LSTM achieves the best performance in this measurement with a sensitivity and precision of 43.4% and 35.2%. The other methods perform considerably worse with both sensitivity and precision values in the range of 5%. Measurement M10737 originates from the iSCI M set and has a walking speed of 0.6 $\frac{m}{s}$. In Figure 4.5 it is shown, that the subject of M10737 has a mirrored course of marker trajectories compared to a healthy subject, hence, the IC in M10737 is initiated with the toe instead of the heel. For detection of IC, ZENI V performs best with an sensitivity and precision of 80.8% in this measurement. The performance of the other methods is around 65%.

TABLE 4.7: Comparison of **mean time difference** between detected events of selected methods and ground truth. Unit of all values is milliseconds [$ms$]

| **Algorithm** | | **healthy set** | | **iSCI L set** | | **iSCI M set** | | **stroke set** | |
|---|---|---|---|---|---|---|---|---|---|
| | | **IC** | **FC** | **IC** | **FC** | **IC** | **FC** | **IC** | **FC** |
| LSTM | | -12.52 ±6.67 | -11.13 ±5.34 | -10.55 ±4.42 | -18.63 ±7.73 | -12.67 ±6.82 | -4.46 ±10.09 | -9.70 ±7.63 | -15.69 ±8.51 |
| SCHOB | | -44.27 ±3.78 | -15.24 ±8.25 | -28.21 ±9.06 | -17.56 ±6.71 | -31.49 ±19.26 | -38.70 ±10.64 | -20.51 ±13.56 | -22.58 ±6.97 |
| ZENI | C | -4.15 ±18.71 | 25.51 ±20.42 | 23.17 ±14.04 | 28.93 ±6.37 | 6.65 ±30.06 | -14.79 ±19.41 | 22.99 ±20.67 | 20.62 ±13.36 |
| | V | -1.16 ±19.94 | 34.03 ±9.34 | 25.91 ±12.40 | 29.49 ±5.38 | 9.05 ±28.90 | -10.82 ±21.53 | 25.66 ±20.22 | 25.31 ±7.64 |

TABLE 4.8: Comparison of **root mean squared time difference** between detected events of selected methods and ground truth. Unit of all values is milliseconds [$ms$]

| Algorithm | | healthy set IC | healthy set FC | iSCI L set IC | iSCI L set FC | iSCI M set IC | iSCI M set FC | stroke set IC | stroke set FC |
|---|---|---|---|---|---|---|---|---|---|
| LSTM | | 18.18 ±7.47 | 16.01 ±4.57 | 16.38 ±3.94 | 21.17 ±7.36 | 21.61 ±8.10 | 26.83 ±3.56 | 17.83 ±5.10 | 19.75 ±6.62 |
| SCHOB | | 45.18 ±3.32 | 19.27 ±7.85 | 31.87 ±7.96 | 19.75 ±6.30 | 34.48 ±15.71 | 39.71 ±9.74 | 26.22 ±11.06 | 24.71 ±6.68 |
| ZENI | C | 22.24 ±8.52 | 33.56 ±8.83 | 30.62 ±7.49 | 30.61 ±5.35 | 30.74 ±13.92 | 28.64 ±9.27 | 34.53 ±9.72 | 27.66 ±8.10 |
| | V | 23.21 ±8.57 | 35.48 ±8.46 | 32.60 ±6.88 | 30.99 ±4.73 | 29.71 ±14.77 | 28.85 ±4.47 | 36.76 ±8.24 | 27.29 ±6.90 |

An observation of the time differences reveals, that LSTM detects all gait events earlier than indicated the ground truth. On average, events are detected 12 $ms$ earlier, which is equivalent to about one frame. The average of root mean squared time differences is about 20 $ms$. In comparison with the other methods, LSTM has the smallest difference to the ground truth in almost all events. Both methods of ZENI have a smaller difference in IC in the healthy set and in the iSCI M set. SCHOB also detects gait events earlier than the ground truth, but has a considerably higher deviation. It is noticeable, that SCHOB has a smaller time difference than LSTM in FC of the iSCI L set. Between time differences of LSTM, there is less scatter. In the other methods, the results have increased dispersion between the events and are more irregular.

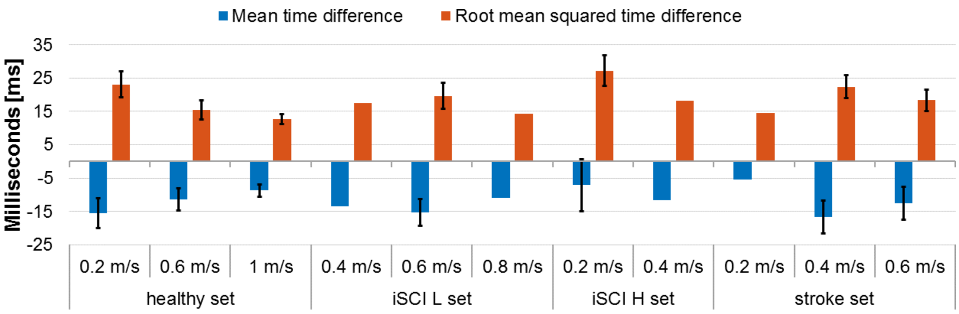


FIGURE 4.6: Averaged mean and root mean squared time difference for detected events of proposed algorithm in relation to different walking speeds of investigated subjects. Bars without error indicator contain only a single measurement of corresponding combination of diagnosis and walking speed

Figure 4.6 shows the timing differences of detected events of LSTM, averaged for IC and FC, in relation to different walking speeds. Timing differences are mostly irregular through the different pathologies and walking speeds. Root mean squared deviation to ground truth is extraordinarily high for 0.2 $\frac{m}{s}$ in the healthy and iSCI M set. Likewise, there is a high spread in the mean time difference of these measurements. A slight trend can be discerned in all sets but iSCI L, where timing differences to ground truth and dispersion decrease with increasing walking speed.

## 4.3 Gait Phase Estimation

This section compares the performance of the selected methods regarding the estimation of gait phases. To compare performance, the mean squared error (MSE) and the mean absolute error (MAE) were calculated and averaged for both body sides. The results of the evaluation are shown in Table 4.9 and Table 4.10.

TABLE 4.9: Comparison of **mean squared error** between estimated gait phases of selected methods and ground truth. All values refer to a percentual gait phase between the values 0 and 1. Unit of all values is percentage [%]

| Algorithm | | healthy set | iSCI L set | iSCI M set | stroke set |
|---|---|---|---|---|---|
| LSTM | P | $1.12 \pm 0.53$ | $1.41 \pm 0.41$ | $2.28 \pm 0.48$ | $1.83 \pm 0.93$ |
| | E | $0.95 \pm 0.56$ | $0.86 \pm 0.32$ | $1.10 \pm 0.66$ | $1.07 \pm 0.51$ |
| SCHOB | | $3.56 \pm 0.88$ | $2.21 \pm 0.96$ | $2.03 \pm 1.22$ | $2.09 \pm 1.67$ |
| ZENI | C | $1.30 \pm 0.91$ | $1.99 \pm 0.82$ | $2.32 \pm 2.06$ | $2.53 \pm 1.14$ |
| | V | $1.43 \pm 1.02$ | $2.19 \pm 0.87$ | $2.50 \pm 2.36$ | $2.82 \pm 1.12$ |

TABLE 4.10: Comparison of **mean absolute error** between estimated gait phases of selected methods and ground truth. All values refer to a percentual gait phase between the values 0 and 1. Unit of all values is percentage [%]

| Algorithm | | healthy set | iSCI L set | iSCI M set | stroke set |
|---|---|---|---|---|---|
| LSTM | P | $6.67 \pm 1.42$ | $6.51 \pm 0.70$ | $9.34 \pm 1.25$ | $8.07 \pm 2.70$ |
| | E | $1.97 \pm 1.23$ | $1.72 \pm 0.64$ | $2.20 \pm 1.33$ | $2.21 \pm 1.05$ |
| SCHOB | | $7.19 \pm 1.85$ | $4.42 \pm 1.93$ | $4.08 \pm 2.46$ | $4.19 \pm 3.36$ |
| ZENI | C | $2.65 \pm 1.96$ | $3.98 \pm 1.65$ | $4.63 \pm 4.13$ | $5.04 \pm 2.31$ |
| | V | $2.93 \pm 2.20$ | $4.41 \pm 1.75$ | $5.02 \pm 4.72$ | $5.65 \pm 2.26$ |

Considering the MSE, performance of all methods differs only slightly from each other. The only noticeable feature is the particularly high error of SCHOB in the healthy set. Distinct differences become visible, when considering the MAE. In comparison, almost all methods have a higher mean absolute error for the clinical set, of which LSTM P has the worst performance of all methods. An observation of the individual measurements showed a systematic shift in the regressed gait phases. While in the implemented methods, this shift is relatively small and settles down occasionally, LSTM P has higher deviations. LSTM E has the best performance in this comparison with a sufficiently small error. Also, LSTM E has no sufficiently large deviations in the metrics of investigated subjects.
An observation of the measurement M10146 showed a drop in performance of all methods. In M10146, LSTM E has the best performance with a MSE of 3.58% and a MAE of 7.46%. This differs by nearly factor four to the achieved performance for healthy subjects. Performance of other methods is around 6.80% for MSE and 14.27% for MAE. In M10737, no decrease in performance could be detected. Here, ZENI C has the best performance with a MSE of 1.16% and a MAE of 2.39%. Performance of the other methods is slightly lower with an average MSE of 1.78% and an average MAE of 3.69%.

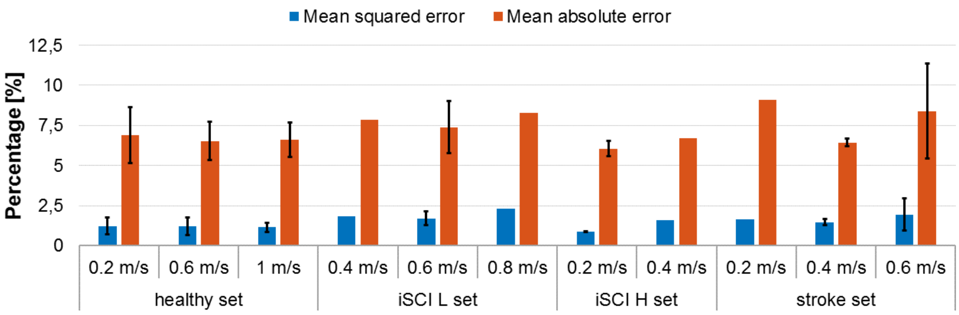


FIGURE 4.7: Comparison of mean squared error and mean absolute error of proposed algorithm for estimation of gait phases via **regression (LSTM P)** regarding different walking speeds of investigated subjects. Bars without error indicator contain only a single measurement of corresponding combination of diagnosis and walking speed

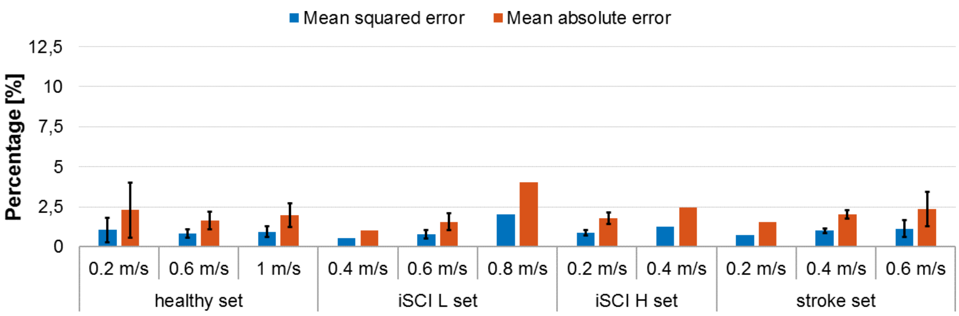


FIGURE 4.8: Comparison of mean squared error and mean absolute error of proposed algorithm for estimation of gait phases via **event detection (LSTM E)** regarding different walking speeds of investigated subjects. Bars without error indicator contain only a single measurement of corresponding combination of diagnosis and walking speed

Figure 4.7 and Figure 4.8 show the mean squared error and the mean absolute error of estimated gait phases of LSTM P and LSTM E in relation to different walking speeds. As before, both metrics and their deviation are considerably lower for LSTM E than. For MAE, this difference is nearly factor three. Deviation for $0.2\,\frac{m}{s}$ in the healthy set is extraordinarily high for both LSTM P and LSTM E. A slight trend can be discerned in all sets, where performance decreases for slow walking speeds, regardless of the subject's diagnosis. For the most part, the values for LSTM P and LSTM E are irregular across all measurements.

# 5 Discussion

This chapter discusses results of the manual annotation and the evaluation with implemented methods. Strengths and limitations of the proposed algorithm are investigated through comparison with existing methods. The choice of approach, achieved requirements and suitability for the feedback system of *SensSCI* is also discussed.

## 5.1 Rater Reliability

In rater reliability, almost perfect agreement was achieved in all methods. Nearly all correlation coefficients yielded *perfect* or *excellent* agreement and low temporal differences were found in the BA analysis and between annotated events among raters. Temporal differences were almost entirely below the considered tolerance of $50\,ms$. The overall results of the used methods for determining agreement show, that a combination of raters is very well suited to determine an unknown ground truth from a data set. Although this approach is time-consuming and tedious, there are some advantages over established methods, like force platforms and footswitches. On one hand, manual annotation can be scaled arbitrarily regarding the granularity of the gait phases and gait events. Furthermore, no external equipment is necessary, which may be a criterion for some laboratories in terms of cost. For determining a ground truth, it is assumed, that this approach is one of the most precise and flexible. Nevertheless, it is beneficial to automate this approach, since the time investment is considerable. A rater took an average $8\,min\,26\,s$ to annotate a single measurement of the subset with the self-developed script. Over all raters and observations, almost $39\,h$ were spent for manual annotation. The manual annotation of the complete set in *Cortex* is not included, as time was not measured during annotation with *Cortex*. Furthermore, the determined ground truth can still deviate from the true ground truth, as all manual processes are exposed to the subjectivity of humans.
In the evaluation of the manual annotation it has been observed, that events of *Final Contact* (FC) have a higher agreement among raters than events of *Intial Contact* (IC). It also became apparent, that raters agree more, when the participant of the observed measurements has a fast walking speed. Reason for higher agreement of FC is, that the termination of foot-floor contact was easier to determine during annotation with the self-developed script than the beginning of it. To annotate gait events as accurate as possible, raters had to determine foot-floor contact of the real foot. This presented one difficulty during annotation, as the real foot of the subject was located lower than indicated by the marker trajectories. Toes of the subjects were also not represented by markers and were located further outside, which raters had to pay attention to. While even under such circumstances a repeating gait pattern could be recognized for healthy subjects, gait patterns of subjects with iSCI or of stroke survivors were more irregular and events more difficult to determine through lack of clarity. This again demonstrated the need to automate this process. A further difficulty was the representation of the floor. The axis, which represents the floor was calculated during an initial calibration prior recording. Because of small deviations between calibrations, the floor axis was not uniform for all measurements. Thus, the IC of the real foot was far below the actual floor axis in some

measurements. Mainly, the events had to be determined by prominent movements of the foot instead of relative positions to the ground. Movements, that initiate FC were easier to identify than those of IC, justifying higher agreement for FC. Another difficulty during annotation were measurements with slow walking speeds. Due to the slower speed, changes in marker positions could not be determined between two frames as clearly as in faster speeds. Therefore, ambiguity prevailed in raters for the exact determination of IC and FC. Frames around possible events had to be examined multiple times, which also consumed more time. For measurements with a walking speed of 0.6 $\frac{m}{s}$ or 1 $\frac{m}{s}$, transitions between two frames were clearly visible. Prominent movements of the foot were easier to determine, which is the reason for the higher agreement.
For IC in healthy subjects, agreement decreases considerably during the second iteration. An observation of time differences between annotated events shows, that the reduced agreement is ascribed to *Rater A*. Timing differences between *Rater A* and the other raters increase considerably in the second iteration. The cause is assumed to be a change in personal perception of *Rater A*, as a reduction in agreement is not found in the other subsets. Since the deviation is significant, it reduces the reliability of *Rater A*. To remedy this circumstance, one possibility would be to have *Rater A* annotate the subset again. If the deviation is no longer present, disagreement could also be explained by a disruptive environment during annotation. However, due to lack of time, a repeated annotation of the healthy set was not conducted. It was also noted that inter-rater agreement for IC in iSCI M increased considerably compared to the other sets. As before, it is assumed, that the reason for this is a change in personal perception of the rater, indicating a learning curve. By gaining experience in the first iteration, raters were more confident in annotating subjects from the clinical set.
Noticeable results are also the significant time differences in intra-rater agreement in IC of the iSCI M set and FC of the stroke set. An examination of the individual measurements revealed, that the deviations in IC of iSCI M are systematic, as every measurement differs around 250 $ms$ from the initial annotation of the complete set, justifying the low scores in *Cohen's kappa*. All subjects of these measurement initiate IC exclusively with their toe instead of their heel. As this constitutes a difficulty in manual annotation, this deviation from a healthy gait pattern is assumed to be the reason for lower agreement. Although, the time deviation for FC in the stroke set is also high, its *Cohen's kappa* is unaffected. Observing the measurements in the stroke set revealed, that the high time differences of FC in the stroke set are caused by a single measurement with an average difference of 800 $ms$. However, the subject of this particular measurement has a relatively healthy gait with only slight pathological characteristics. While the deviation in time could be justified by the personal perception of *Rater A*, it is assumed, that the reason for the high deviation is a change in the annotation method, as the annotation of the CS took place in *Cortex*, while the subsets were annotated using a self-developed script. An important difference between these two methods is the viewed perspective. In *Cortex*, the perspective is freely selectable in a three-dimensional space. In the self-developed script, the view is fixed, displaying only movements in anterior-posterior direction. Since this deviation only exists in intra-rater agreement, it is assumed that a wrongly set perspective in *Cortex* is the cause of it. Due to an unsuitable perspective during one annotation session, the rater was biased and determined events at wrong time instances.
Another conspicuity in the results of manual annotation are considerably higher scores for ICC than for *Fleiss' kappa*, although both determined inter-rater agreement. An observation of data processing and given notes within the works of Cardillo G. [28, 29] and Arash S. [74] showed, that data within *Fleiss' kappa* is processed as ordinal data while data within ICC is processed as continuous data. Since the annotated events are instances of time, therefore continuous data, ICC is more suitable for processing and

thus provides better scores. It is assumed that converting events into categories prior processing by *Fleiss' kappa* distorts the contained information of event times. Since the events are processed as categories, the arrangement of data is also ignored. This hinders an appropriate assessment of rater agreement with the given data, making *Fleiss' kappa* unsuitable for determining rater agreement of continuous data. As both *Cohen's kappa* and *Fleiss' kappa* process ordinal data, it is not recommended to use them for determining the level of agreement between different annotations of gait events. In this thesis, it would have been appropriate to determine intra-rater agreement using ICC as an addition to *Cohen's kappa* in order to obtain results of nominal data processing. In literature [14], the use of kappa is also criticized as it assumes independent ratings when measuring the level of agreement. If several individuals rate the same observation, the agreement among raters will be overestimated. This criticism can not be applied in this thesis, as ICC was calculated as an alternative to *Fleiss' kappa* and ICC scores are considerably higher than the supposedly overestimated scores of *Fleiss' kappa*. In comparison with existing methods, only three approaches [66, 33, 47] were found, in which gait data was manually annotated. In other methods [23, 35, 41, 42, 43, 69, 70], mostly the force plate method is used to create a reference. Two of the found approaches [66, 47] with manual annotation also use ICC to determine rater agreement. In these works, excellent results were achieved for rater agreement via ICC, too. Therefore, the use of ICC to determine agreement in annotated gait events is recommended. A disadvantage of using ICC is, that one has to select the appropriate model. Selecting the wrong model will skew the results for the corresponding case of application. Nonetheless, guidelines [52] exist to support the selection of appropriate ICC and their use is recommended.
Regarding Bland-Altman analysis, it is increasingly used in gait analysis [33, 66, 17, 72]. Mainly, it is used to evaluate a proposed method with a reference. However, this thesis and the work of Mickelborough et al. [66] show, that it can also be used to determine agreement among raters for manual event detection. Without the use of BA plots, it would not have been possible to detect the potential connection between rater agreement and walking speed. As a large amount of data was mapped in each plot, determining general agreement proved to be difficult. It is recommended to use BA plots with a smaller number of data points. Looking at the literature of existing methods [66, 33, 47], it was noticed that often only agreement regarding the subject's diagnosis is considered, and not agreement regarding walking speed. Since the results of this thesis indicate a connection between rater agreement and walking speed, it is recommended to observe this in future works.
Despite a change in annotation methods for *Rater A* and high discrepancies in the second iteration of annotation, mean deviation between raters is still less than the considered tolerance. Scores of correlation coefficients and results of BA analysis are almost perfect, too. This approach is therefore suitable for determining an unknown ground truth. All in all, the combined use of BA analysis and correlation coefficients was considered optimal in this thesis and is recommended for future works. Regarding correlations coefficients, there is no assessment of the confidence intervals in this thesis, as it was considered to be not in the scope of this thesis. However, this shall be made up for within a possibly planned publication.

## 5.2 Event Detection

In comparison with the implemented methods of ZENI and SCHOB, the proposed algorithm proved to be superior in terms of all considered metrics. The spread of the metrics is also reliably small. Thus, LSTM can be classified as the best algorithm in this comparison in terms of event detection. These results indicate, that better performance can be achieved with a data-driven processing using neural networks. However, simpler processing has the advantage of a lower computational load and thus a lower processing time. An observation of recorded gait data showed, that average length of a stride in a non-disabled subject is around 100 frames. With a sample rate of 60Hz, a stride has a length of 1600 $ms$. Using an Intel i5-4760K CPU (3.4GHz), processing time in the proposed algorithm is $42.29 \pm 2.76\,ms$ per frame, which is under 1% of the total duration of the stride. For an use in *SensSCI*, the processing time of the feedback system has to be taken into account, too. As was shown in the results, the proposed algorithm detects events on average earlier than indicated in the ground truth. This supports the processing in a connected application, as a feedback system needs also some amount of time to process data.

An observation of individual measurements revealed, that performance for detection of events is mostly independent from walking speed or the diagnosis of the corresponding subject. Only for a walking speed of $0.2\,\frac{m}{s}$ a drop in performance could be detected. This indicates robustness and a sufficient generalization of the algorithms for detection of events in investigated subjects. The observation revealed, that temporal differences are also independent of the subject's diagnosis. Temporal differences are higher if the corresponding measurement has a slow walking speed. This decrease of performance also occurs in the recording of wearable sensors [50, 78, 27, 38, 72] and is an universal problem in the field of gait analysis. It should be noted, that gait speeds of $0.2\,\frac{m}{s}$ and $0.4\,\frac{m}{s}$ are very slow compared to the average speed of non-disabled subjects and can only be found in individuals with gait pathologies. It is assumed, that the slower walking speed causes the gait pattern to change sufficiently to hinder the algorithms in detecting prominent changes or arrangements of values, thus limiting the performance. This is considered to be a limitation of the proposed algorithm.

Regarding the secondary contribution of this thesis, performance of the self-developed and frequently used algorithm of the *Spinal Cord Injury Center* can no longer be called contemporary. The results of the event detection showed inferiority both to the proposed algorithm and ZENI. Although time differences of SCHOB are on average slightly smaller than of ZENI, differences are still considerably higher than in the proposed algorithm. In addition, SCHOB is clearly inferior to both methods in terms of sensitivity and precision. As already stated in the aforementioned requirements, for a reliable use in *SensSCI*, the considered algorithm must not only detect all possible events, but also have as little temporal difference as possible. Both requirements are not met by SCHOB, therefore, a use is not recommended. It is suggested to use the proposed algorithm instead, since it fulfills the requirements with the best performance in this comparison.

Within the results, measurements M10146 and M10737 showed a significant drop in performance for detection of IC. A detailed examination of M10146 showed slight deviations from the gait pattern of a non-disabled subject. In each stride, the subject's feet are swung slightly further before initiating for IC. Although no considerably deviations can be identified visually, this represents a limitation for the algorithm. In M10737, the subject has a very irregular pattern for IC. Strides are initiated with the entire foot or with the toe by the subject. While this presents also a limitation for the proposed algorithm, the drop in performance occurs only in this particular measurement. Other measurements from subjects with a similar gait pattern are not affected and are successful examples for a generalized event detection of the proposed algorithm. One approach to overcome this limitation would be the use of force plates for the detection of IC. This is also an example, that the mere use of kinematic data can limit event detection. It should also be noted, that the achieved performance in evaluation metrics is higher for M10737 than for M10146, although M10146 originates from the healthy set. A further swung foot represents a higher deviation from the learned gait pattern than a pathological gait in the proposed algorithm. This could be an indication of insufficient generalization, as it shows that slight forgeries in gait pattern can considerably reduce performance. As all subjects from the clinical set are classified as well walking subjects, a drop in performance is anticipated for several impaired individuals or individuals with gait pathologies beyond the trained scope. It should be noted, that the investigated clinical set contains only 23 measurements of two different pathologies. It needs to be examined to what extent this represents a potential limitation. However, as the performance of neural networks is highly depended on available training data, the presented limitations could be overcome with sufficient inclusion of the particular types of gait pattern in training data.

For further evaluation of the performance of the proposed algorithm, published results of existing gait detection methods were compared with the results of this thesis. Table 5.1 presents the comparison between the proposed algorithm and existing gait event detection methods. All methods detect events in optical marker-based systems during overground or treadmill walking. Due to different published metrics, the absolute time difference was additionally calculated for the proposed algorithm. Results of subjects with pathological gait were merged in all methods to allow an easier comparison.
As in the comparison with implemented methods, the proposed algorithm detects events earlier than indicated in the ground truth with consistent small deviation, regardless of the walking speed or diagnosis of the viewed subject. This is in contrast to existing methods, like Hansen et al. [35] or Desailly et al. [23], as there are significant time differences between detection of IC and FC. Considering time differences for both events, the proposed algorithm proves to be superior to Desailly et al. [23], Hansen et al. [35], Hsue et al. [43], Kidziński et al. [51] and O'Connor et al. [69]. Methods of Hreljac et al. [42, 41], Miller [67] and Zeni et al. [94] show smaller time differences, but are overall similar in magnitude. Regarding coverage, the proposed algorithm covers 84.49% of all events of the ground truth within $16.67\,ms$, 94.93% within $33.34\,ms$ and 97.05% in a scope of $50\,ms$. Thus, the performance of the proposed algorithm is slightly inferior to the methods of Ghoussayni et al. [33] and Osis et al. [70]. All methods but Miller [67] and Kidziński et al. [51] rely on simple data processing via heuristics. It should be noted that the published results are not consistent in used metrics and were carried out under different conditions. A comparison is therefore difficult to conduct. The superiority of the named existing methods can not be determined exactly. Most noticeably in the case of Zeni et al. [94], whose method is superior according to the literature, but underperforms in the domain of the proposed algorithm. Also, all superior methods are designed for the offline detection of gait events. It must be investigated, whether these methods can even be adapted for real-time, as this adaptation represents a further challenge in itself. Regarding rate of event detection and processing time, it is possible, that some of the superior methods are not suitable for real-time detection at all. Of the presented existing methods, only three methods [33, 51, 70] published the coverage of the detected events. This fact is to be criticized, because it is unlikely for an algorithm to correctly classify all given samples, including false negative and false positive. An algorithm can not be considered suitable for an application, if the temporal difference of detected events is very small, but only a fraction of events are detected at all.
An observation of the experimental protocol shows significant differences in used subjects, as some existing methods [23, 43, 69] tended to test gait event detection on children as well. The detection of gait events in children was no in the scope of this thesis. No comparison can be made regarding this aspect. It is assumed, that children, as their leg muscles are not fully developed yet, have a slower walking speed. Results of the proposed algorithm showed a limitation on slower walking speeds. Thus it is assumed,

that the proposed algorithm will perform poor on event detection in children. However, to determine this, data for evaluation is required.

TABLE 5.1: Comparison of existing gait event detection methods with proposed algorithm. As Hsue et al. and Zeni et al. propose two methods each, a further distinction is done. Methods of Hsue et al. are abbreviated as AP (anterior-posterior) and V (vertical). Methods of Zeni et al. are abbreviated as C (coordinate-based) and V (velocity-based)

| **Author** | | **Subjects** | **Performance** [$ms$] **IC** | **Performance** [$ms$] **FC** | **Metric** | **Real-Time** |
|---|---|---|---|---|---|---|
| Proposed algorithm | | healthy subjects | $-12.52 \pm 6.67$ | $-11.13 \pm 5.34$ | mean time difference | Yes |
| | | pathological subjects | $-10.45 \pm 6.46$ | $-12.36 \pm 9.59$ | | |
| | | healthy subjects | $14.05 \pm 7.28$ | $12.81 \pm 4.98$ | absolute time difference | |
| | | pathological subjects | $13.67 \pm 5.44$ | $15.33 \pm 7.28$ | | |
| Desailly et al. [23] | | healthy subjects | $27 \pm 19$ | $-14 \pm 12$ | mean time difference | No |
| | | pathological children | $1 \pm 23$ | $-2 \pm 25$ | | |
| | | healthy subjects | $29 \pm 17$ | $17 \pm 7$ | absolute time difference | |
| | | pathological children | $17 \pm 15$ | $18 \pm 17$ | | |
| Ghoussayni et al. [33] | | healthy subjects | 90% of all events within 16.7 $ms$ | | absolute time difference | No |
| Hansen et al. [35] | | healthy subjects | $7.5 \pm 5.83$ | $18.08 \pm 9.83$ | mean time difference | No |
| Hreljac et al. | [41] | healthy subjects | 1.2 | 1.2 | mean time difference | No |
| | [42] | | 2.4 | 2.8 | | |
| Hsue et al. [43] | AP | pathologial children | 9 | 20 | absolute time difference | No |
| | V | | 15 | 25 | | |
| Kidziński et al. [51] | | pathological children | 18.3 | 12.5 | mean time difference | Yes |
| Miller [67] | | pathological subjects | $-3.96 \pm 13$ | $-6.05 \pm 14.5$ | mean time difference | No |
| O'Connor et al. [69] | | healthy children | $16 \pm 15$ | $9 \pm 15$ | mean time difference | No |
| | | pathological children | $-3 \pm 9$ | $-6 \pm 26$ | | |
| | | healthy children | 15 | 11 | absolute time difference | |
| | | pathological children | $7 \pm 6$ | $23 \pm 10$ | | |
| Osis et al. [70] | | not named | 89-94% of all events within 20 $ms$ | | absolute time difference | No |
| Zeni et al. [94] | C | healthy subjects | -16.2 | 5.7 | mean time difference | No |
| | | pathological subjects | 20.45 | -6.9 | | |
| | V | healthy subjects | -5.2 | 3 | | |
| | | pathological subjects | 0.25 | 1.75 | | |

The experimental protocol also shows differences in the number of used cameras and sample rate. Some approaches [42, 50, 41, 69] use less than four cameras for recording marker trajectories. As more cameras capture more room volume, the occurrence of gaps in the recording of trajectories, for example when a marker is obscured by an obstacle, is less likely with an increasing number of cameras. Despite good performance, it is assumed, that approaches like Hreljac et al. [42] may not provide a reliable use in feedback systems like *SensSCI*, as a seamless recording of data is required. Regarding sample rate, most of existing methods [67, 69], that are superior to the proposed algorithm, have a sample rate of 120Hz. In a real-time event detection, sample rate is an important factor as each frame is processed individually. In the proposed algorithm, temporal differences are approximately $12\,ms$ with a sample rate of 60Hz. If the sample rate is set to 120Hz, temporal differences may be reduced, assuming that the difference in frames remains the same. Superiority to the methods of Miller [67], Ghoussayni et al. [33] and Osis et al. [70] could be achieved, making the proposed algorithm the best method in this comparison. However, further examination is needed.
Another difference in the experimental protocol is the walking speed. As already stated, the proposed algorithm has lower performance for slower walking speeds. In comparison with existing methods, it is noticeable, that some approaches [67, 51, 70] have considerably faster speeds in their protocol. Despite gait pathologies, Miller [67] and Kidziński et al. [51] have subjects with walking speeds faster than $0.8\,\frac{m}{s}$. It is assumed, that these subjects have less restrictions in the locomotor system compared to the investigated subjects. Thus, this may contribute to the good performance of Miller [67] and Kidziński et al. [51]. Above all, Osis et al. [70] stands out with subjects with gait speeds of $2.4\,\frac{m}{s}$ to $2.8\,\frac{m}{s}$, which can be classified as fast jogging or running. As this type of movement was not in the scope of this thesis, it was not examined. It is assumed, that the proposed algorithm will fail at such speeds, as there are clear differences in gait pattern in comparison with the trained pattern. As such high speeds are not intended in the feedback system of *SensSCI*, this limitation can be ignored. However, it represents a possible extension of the algorithm for a more overall generalization in the detection of gait events.
The only method, that makes a direct comparison possible, is of Kidziński et al. [51], as it also uses LSTM RNNs for online event detection. Coverage of Kidziński et al. [51] was 99% for IC and 95% for FC. The proposed algorithm has a coverage of 95.6% for IC and 98.5% for FC. An approximately equal number of events is discovered by the proposed algorithm with lower time difference than the approach of Kidziński et al. [51]. However, Kidziński et al. [51] investigated children with different pathologies at gait speeds of $0.56\,\frac{m}{s}$ to $1.14\,\frac{m}{s}$. In contrast, this thesis examined gait speeds of less than $0.4\,\frac{m}{s}$, which is considered to be a more difficult condition.
With high detection rate and low temporal difference, the proposed algorithm provides a remarkable performance for the intended use. It is concluded, that the performance of the proposed algorithm is sufficient for most clinical and research applications, even in real-time. In comparison with existing methods, this is considered to be a unique feature. In terms of temporal differences, the proposed algorithm has no clear superiority over existing methods. As stated, a comparison is difficult. It has been shown that more complex data processing does not guarantee better performance. Simple conventional methods can offer a similar or even better performance, when the differences in experimental protocols are ignored. However, conventional methods are usually algorithms for offline detection of gait events, whose complexity will increase if they have to be adapted for real-time detection. In this case, the proposed algorithm has more potential. It is assumed, that the selected features and configuration are sub-optimal, which offers room for optimization. With more training data, changed architecture and configuration it is assumed, that a much higher performance can be achieved and thus general superiority

of the proposed algorithm. Ultimately, algorithms for gait event detection remain limited to their domain, and an optimal solution may need to be sought outside of the presented methods. With the possibility for high detection rate and generalization, it is assumed, that the use of neural networks will considerably improve the current state-of-the-art.

## 5.3 Phase Estimation

A comparison of the implemented method displays, that the proposed algorithm has the worst performance for phase estimation via regression. If the gait phase is estimated from time intervals between detected events instead, the proposed algorithm provides the best results. Since the proposed algorithm achieves the best results in gait event detection, the results can be transferred directly to this comparison. This suggests, that a regression of gait phases is inferior to a estimation via time differences between detected events. It should be noted, that LSTM P was the only method evaluated under real-time conditions. For a comparison with LSTM E under real-time conditions, it is assumed, that the performance of LSTM E will be lower than in the presented results. However, as LSTM E and LSTM P differ in performance by almost factor three, it is assumed, that LSTM E will still outperform LSTM P in an online comparison. The estimation of gait phases via event detection has the disadvantage, that it is based on an heuristic prediction of future events via time intervals between detected events. For a reliable estimation, all events have to be detected seamlessly, which is not always guaranteed. If only one event is missed, the calculation of the estimation will be distorted. A complex post-processing is required to prevent this, which increases computational load. The advantage of a data-driven estimation, as shown in LSTM P, is, that the phase regression is completely independent of detected gait events. The gait phase is calculated directly from a data input and no prediction for the future has to be made. Processing within LSTM P also consumes less computation load since no post-processing is required. Although LSTM P is clearly inferior to LSTM E, its use in applications is not excluded. LSTM P can still be used in applications with low requirements, for example, determining a time window for electrical stimulation of individuals. For applications with higher requirements, like controlling an exoskeleton, the performance of LSTM P is insufficient. As already stated, a change in selected features and configuration of neural networks offers room for optimization. It would have also been appropriate to implement a popular method for real-time estimation of gait phases in this thesis in order to enable a comparison within this domain. As no suitable method could be found in the course of this thesis, this opens the possibility for further investigations in future works.
As in the detection of events, a drop in performance was found for slower walking speeds. This is, as already stated, an universal problem for applications in gait analysis. For all methods in the estimation of gait phases, a decrease in performance was also found in measurement M10146. A further swung of the leg thus represents a limitation in the proposed algorithm for estimation of gait phases, too. However, conducting IC with flat-foot or toe does not limit the performance of the proposed algorithm in phase estimation, indicating an overall generalization.

Table 5.2 compares the results of the proposed algorithm with the published results of existing gait phase estimation methods. Due to different published metrics, the root mean squared error was additionally calculated for the proposed algorithm. Results for subjects with pathological gait were averaged in all methods to allow an easier comparison.

Table 5.2: Comparison of existing gait phase estimation methods with proposed algorithm. As all presented methods estimate a percentual gait phase between the values 0 and 1, all results were converted into percentages to allow an easier comparison

| **Author** | **Subjects** | **Performance [%]** | **Metric** | **Real-Time** |
|---|---|---|---|---|
| Proposed algorithm (LSTM P) | healthy subjects | $1.12 \pm 0.53$ | mean squared error | Yes |
| | pathological subjects | $1.84 \pm 0.61$ | | |
| | healthy subjects | $6.67 \pm 1.42$ | mean absolute error | |
| | pathological subjects | $7.97 \pm 1.55$ | | |
| | healthy subjects | $10.7 \pm 2.2$ | root mean squared error | |
| | pathological subjects | $13.2 \pm 2.2$ | | |
| Proposed algorithm (LSTM E) | healthy subjects | $0.95 \pm 0.56$ | mean squared error | No |
| | pathological subjects | $1.01 \pm 0.49$ | | |
| | healthy subjects | $1.97 \pm 1.23$ | mean absolute error | |
| | pathological subjects | $2.04 \pm 1.01$ | | |
| | healthy subjects | $9.3 \pm 2.5$ | root mean squared error | |
| | pathological subjects | $9.6 \pm 2.4$ | | |
| Jiang et al. [47] | healthy subjects | 2.1 | averaged difference | No |
| Mariani et al. [58] | healthy and pathological subjects | $0.6 \pm 1.5$ | averaged difference | No |
| Senanayake et al. [76] | healthy subjects | $1.26 \pm 6.08$ | averaged difference | Yes |
| Trojaniello et al. [83] | healthy subjects | 5.5 | mean absolute error | No |
| | pathological subjects | 5.3 | | |
| Vu et al. [86] | healthy subjects | $0.8 \pm 0.1$ | mean squared error | No |
| | | $3.2 \pm 0.2$ | mean absolute error | |
| Yan et al. [92] | healthy subjects | 1.1 | root mean squared error | Yes |
| Zheng et al. [95] | healthy subjects | 4.1 | root mean squared error | Yes |

As in the results of the detection of gait events, a direct comparison is difficult to conduct, since in addition to different experimental protocol and subjects, existing methods for the estimation of gait phases also use different sensors to record gait data. None of the presented methods uses an optical marker-based system for estimation of gait phases.
Regarding the comparison with other existing methods, the performance of LSTM P is mediocre, and no superiority can be stated. All other existing methods show better performance in estimation of gait phases. Especially the superiority of Zheng et al. [95] becomes clear, since the values of the corresponding metric differ by factor ten. However, LSTM P splits off from the presented methods, as, according to conducted research of literature within this thesis, it is the first method, that calculates the percentual gait

phase directly in real-time. As shown in the results, a much better performance is achieved when computing gait phases with detected events. However, this approach is not superior to all presented methods. The methods of Vu et al. [86] and Mariani et al. [58] offer an approximately equal performance to the proposed algorithm. The approach of Senanayake et al. [76] is also real-time capable, but has a slightly lower performance with considerably higher deviation. As all of these methods use IMUs to calculate gait phases, the proposed algorithm offers an advantage over these, as there is no encumbrance of the observed subject. Furthermore, the performance of sensors like IMUs is heavily dependent on the body placement, which also does not apply to the proposed algorithm. It is also important to mention, that LSTM E is real-time capable, but was not evaluated under these conditions. Methods of Yan et al. [92] and Zheng et al. [95] still provide better performance and additionally estimate gait phases in real-time.
It is noticeable, while performance of the proposed algorithm can be classified among IMUs, methods with superior performance use oscillators. In comparison, the use of oscillators shows the best performance of all considered methods. A major difference between oscillators and the proposed algorithm is the different consideration of gait data. While in the proposed algorithm gait is considered as a time series, in oscillators the gait is considered as a periodic signal. Both methods [92, 95] forecast the whole gait percentage of a stride and adapt their respective model through phase error compensation. While this approach offers outstanding prediction, several parameter have to be finely tuned by trial and error. Likewise, prediction performance drops during speed transitions and free walking. It should also be noted, that Yan et al. [92] and Zheng et al. [95] have evaluated their methods only with healthy subjects.
Overall, approaches with oscillators are ranked as having the best performance in this comparison, followed by methods using IMUs. If gait phases are estimated via detected events, the performance of the proposed algorithm can be classified among methods using IMUs. The phase regression of the proposed algorithm is inferior and has a moderate performance, indicating that the use of long short term memory recurrent neural networks may not be suitable. A use is not recommended. Instead, it is recommended to calculate the gait phase using time differences of detected events.

## 5.4 Choice for Neural Networks

For the proposed algorithm, long short-term memory recurrent neural networks (LSTM RNNs) were used. Rather than relying on simple data processing as in conventional approaches, it was assumed to achieve better performance with more sophisticated processing. For this, the stride was considered as a time series of interdependent values. Results have shown, that conventional approaches can often achieve the same performance as the proposed algorithm and thus refute the assumption. However, most of conventional methods refrain from real-time processing. All considered methods, that are capable of real-time processing or stated to be [51, 67, 86, 92, 95], use a data-driven processing, even with neural networks [51, 67, 86]. Including the proposed algorithm, neural networks prove to be an accurate and performant solution for problems of gait analysis. The expectation is, that, as technology advances, neural networks will be used excessively in the area of gait analysis.
The use of time series for detection of gait events has already been investigated by other authors. Mannini et al. [57] uses hidden Markov models [6] in combination with the Viterbi algorithm [85] to detect four gait events. This method achieved a good performance on healthy subjects but had a high standard deviation in time differences. Due to this irregularity and the fact, that performance often decreases for individuals with gait

pathologies, it was assumed, that the approach of Mannini et al. [57] would not generalize sufficiently for the investigated data set and would therefore be unsuitable for use in a feedback system of *SensSCI*. In order to choose a flexible approach, neural networks were selected as a possibility. Finally, to eliminate the described disadvantages of neural networks regarding temporal dependencies, LSTM RNNs were chosen as the approach for this thesis.

In retrospect, it turned out to be a good choice. The approach is suitable for the presented case of application and provides high performance for the chosen evaluation metrics. Regarding detection rate and temporal differences in gait event detection, results are also generalized well for different gait speeds of investigated subjects. There are only minor differences in performance between the healthy set and the clinical set. The processing time of $42.29 \pm 2.76\,ms$ is also low enough for reliable use in real-time applications. However, the processing time represents the processing time of a single LSTM cell. As there are four LSTM RNNs in the proposed algorithm, the total processing time would be around $160\,ms$ in a single-core application. Depending on the application, the tolerable time scope would be exceeded. It is therefore necessary to use parallel LSTM RNN cells in a multi-core application. Furthermore, the proposed algorithm can be scaled arbitrarily regarding event detection. The LSTM RNNs are capable of detecting multiple types of gait events. For this purpose, only the training data must be adapted. With regards to the regression of gait phases, only moderate performance was achieved. It is assumed that, this is not related to the inability of the network, but to the training parameters and architecture. This represents also a disadvantage of this approach, as for good performance intensive and heavy parameter tuning must be conducted. If one has not the necessary expertise, getting suitable results is laborious. Likewise, due to the characteristics of neural networks, there is no way to comprehend the distribution of the weighting within the network. This behavior, which resembles a black box, precludes the optimization of already trained networks and is considered to be a possible disadvantage of using neural networks.

All in all, neural networks offer a good approach for the problems of gait analysis. Within the results of the event detection and phase regression, it emerged, that time series classification via LSTM RNNs is exceptional well suited for detection of gait events. On the other hand, time series regression using LSTM RNNs is only moderately suitable for estimation of gait percantages.

An alternative and novel approach for classifying time series is presented in Wang et al. [87], Cui et al. [21] and Karim et al. [49]. So-called *fully convolutional neural networks* (FCNNs) are used, which provided very good performance in comparison with other state-of-the-art methods for the classification of time series. This approach was also tried out in this thesis, but could not be applied to the posed problem. The networks were not able to learn in both event detection and phase regression. Due to time constraints and the fact, that this approach did not go beyond the first steps, this approach was discontinued. However, it is assumed, that these cases of application can also be carried out with a suitable configuration of FCNNs.

Another alternative has already been presented in the methods of Yan et al. [92] and Zheng et al. [95], in which the stride is considered to be a periodic signal. At each time step, an error signal is calculated to determine the difference between estimated and measured signal. This error signal is used to update the adaptable parameters for the estimation of gait phases in the stride cycle. Results of Yan et al. [92] and Zheng et al. [95] have shown an outstanding performance in estimation of gait phases. A disadvantage is, that several sensitive parameters have to be tuned by trial and error prior processing and for each subject individually. The method also takes some time to adapt to the new signal when walking speed changes. Furthermore, only healthy subjects were examined by Yan et al. [92] and Zheng et al. [95]. It is therefore uncertain, whether the good results can also be transferred to individuals with gait pathologies. However, this approach can be well combined with the proposed algorithm to improve the moderate performance in phase estimation using regression. For each IC, the proposed algorithm can predict the phase progression of the entire stride and use the times of FC and next IC as milestones for the 60% and 100% marks of the phase progression, respectively. As in the presented approaches, the model parameters would be updated by a calculated error signal at each occurrence of a milestone. It is assumed, that, regarding the good performance not only in the presented methods [92, 95], but also in other elaborations [88], this approach can achieve the best possible estimation of gait phases in the proposed algorithm and should be considered for future works.

## 5.5 Architecture and Parameter Selection

The selected features for the proposed algorithm are shown in Table 3.2 and Table 3.3. For the selected features, it should be noted, that there are no features containing joint angles. Mainly velocity and acceleration data were selected within the feature selection, as these, according to random forest, have the highest feature importance. Although the found features coincide with features of existing methods [86, 69, 84, 71, 43, 94, 33], joint angles are also used repeatedly in published works [42, 57, 70, 58, 76]. A reason for the absence of joint angles could not be found besides the chosen feature selection method via random forest. It would have been appropriate to perform another feature selection method to evaluate the results of the random forest. One possible approach would have been to use the LSTM RNN as a feature selection method by training the LSTM RNN with all extracted features and remove features without given weights afterwards. However, such an approach is not recommended, as it consumes a considerable amount of time. Even with the smallest feature set, the network had a training period of several hours. Yet, results of evaluation show good performance with the selected features. In comparison with other neural networks for the detection of gait events, the feature set used in this thesis is also significantly smaller. The approach of Miller [67] needs 15 features on two bodies to detect gait events on a single leg, while Kidzinski et al. [51] track 33 features on five bodies. The proposed algorithm observers only 8 features on three bodies for classification, thus needs less input for providing equal or better performance in event detection. As already stated, it is assumed, that the performance of the proposed algorithm could be further improved with usage of a different feature set, both in event detection and phase regression.
Regarding data filtering, a Butterworth filter was used in the proposed algorithm. For real-time processing, data has to be filtered unidirectional, which leads to a phase shift of five frames. Despite the phase shift, good performance is achieved by the proposed algorithm and processing is not largely hindered. As the raw data shows only relatively low distortion and most noise is located between prominent areas, the question arises whether filtering is necessary at all. Removing the filter could have the disadvantage of reducing the detection rate of the proposed algorithm. However, it would greatly reduce the temporal differences for gait event detection. To what extent the advantages outweigh the disadvantages must be further examined.
The improvement of performance can be also applied to the architecture of the LSTM RNNs. Iteratively, various training parameters and architectures were tried out, of which ultimately those with the best performance were chosen. However, one change in configuration may not work well in itself, but with other selected changes, constituting a multi-dimensional optimization problem. Although some parameters and architectures

could be excluded in the course of the development, the development process had to be discontinued due to lack of computing power and time constraints. Thus, some local maxima could be determined during development, but it is uncertain, if the global maximum of the multi-dimensional plane, such as optimal training parameters and architecture, was found. As with the feature selection, performance of the LSTM RNNs can still be increased and the algorithm itself further simplified. It is assumed, that revealed limitations of the proposed algorithm can also be removed with proper configuration.

Finally, the improvement of performance can also be applied to the choice of sequence length. Changes in architecture and training parameters will also require a different sequence length for optimal performance. For the proposed algorithm, the selected sequence length for event detection consists of 12 frames before the actual event and 2 frames after the actual event, while for regression of gait phases the chosen sequence consists of 27 frames before the actual event. The sequence in event detection presents a limitation as at each processing step the second last frame is classified retroactively. It is compensated by the fact, that even with retroactive classification, the time difference to ground truth remains negative. If retroactive classification is removed by post-processing and moved to the current frame, the temporal difference would only increase. As such, no post-processing is needed for the proposed algorithm. Regarding optimization, a change in sequence length would also have an effect on the built-up time of the proposed algorithm, as a change in sequence length also changes the required number of frames for the first classification or regression of the proposed algorithm. Currently, the algorithm requires a built-up of 466.676 $ms$, or approximately 0.45 $s$, to be fully operational. Since each investigated measurement is 90 $s$ long, this one-time delay was considered to be tolerable.

For development, focus was put on the approach of Stathakis et al. [81], which stated, that for solving most problems, a neural network with a single hidden layer is sufficient. Following this, single cell LSTM RNNs were used for each case of application in this thesis. A benefit of using single cells is, that features are not processed as complex as in multiple connected cells. Due to the connected processing, features could be changed during processing to such extent that they no longer resemble the originally designated information and thus are interpreted differently by the network. Furthermore, the simpler architecture uses less computational load, which benefits processing in real-time. Considering the results, remarkable performance can be achieved using a single LSTM RNN cell. During the search for a proper network architecture, the method of Kidzinski et al. [51] was also tried out. Its architecture consists of three connected LSTM cells with 16, 32 and 64 hidden units. Testing the architecture of Kidzinski et al. [51] with the selected features of this thesis revealed, that no better performance could be achieved. Much more, this architecture required more computational load than the single cell approach, as in this case, three cells had to be updated at each time step. For this reason, using multiple cells for each case of application was no longer pursued.

During development, it was also revealed, that performance improves, when event detection focuses on only one side of the body. Reason for this is, that, with enlarged feature input and output classes, data processing becomes more complex. This was investigated by Berstad et al. [9], where a mulitclass classifier competed against multiple binary class classifier. The result was, that splitting the multiclass problem into binary classes increases performance, as a higher number of output nodes results in a higher complexity of processing. However, the splitting leads to a higher resource consumption. In this thesis, for each body side, a LSTM RNN was used for detection of gait events was to get the best trade-off between detection rate and computational load. Alternatives are, using a single LSTM RNN for detection of events on both body sides or using four LSTM RNN for detection of each event of the corresponding body side.

All in all, a remarkable performance was achieved with the chosen training parameters and the selected architecture. The proposed algorithm can keep up with existing methods or even partially outperform them. Through intensive procedures, parameters and architecture were tried out during development. Nevertheless, it is assumed, that the optimal configuration has not been found yet. Such, the performance of the proposed algorithm can be further increased and current limitations removed.

## 5.6 Achieved Requirements

For a reliable use in feedback systems, several requirements were set for the proposed algorithm. One of the requirements was, that the proposed algorithm has to offer the possibility to be integrated into existing systems as easily as possible. This requirement can be considered fulfilled. The proposed algorithm is used as a data processing pipeline. It can therefore be easily integrated into existing applications. However, it should be noted that the trained network was developed for usage in optical marker-based systems and thus is mainly suited for such systems. For a use of the proposed algorithm, the Helen Hayes marker set or a similar marker set, which contains the features described in Table 3.2 and Table 3.3, must be used. Depending on the given processor performance, it is also advisable to run the individual LSTM RNN cells of the algorithm in parallel to save processing time. Due to differences in performance, only the use of the gait event detection is recommended. For the estimation of gait phases, it is recommended to compute the phases out of time distances of detected events. It is planned to revise the regression of gait phases and replace it with another procedure.
Another requirement for the event detection was a high recognition rate with a small delay. This requirement has also been met. The proposed algorithm has a coverage of 97.05% for the detection of events within a tolerance of $50\,ms$. 1.83% of the ground truth was detected, but lies beyond this tolerance. Only 1.12% of the ground truth could not be detected. 58 additional events were found by the proposed algorithm, which are not located in the ground truth. In relation to the ground truth, this amounts to 0.42% of additionally detected events. It is an excellent method for the detection of events in both healthy individuals and individuals with gait pathologies. As revealed in the measurements M10146 and M10737, further swing of the foot prior IC as well as a tread with the toe or entire foot leads to a decrease in detection rate. Since all investigated subjects are characterized as well walking, it can be assumed, that more severely affected individuals of the examined pathologies and individuals with other neurological or orthopedic gait disorders constitute also a limitation for the proposed algorithm. Another limitation is suspected to be in the detection of events in children and running subjects, as their gait patter differ from the trained pattern. However, no data was available to examine this. Since the proposed algorithm was not trained on these gait pattern and these patterns were also not within the scope of this thesis, this is not considered to be a limitation. Lastly, an increase in time differences of detected events occurs for measurements with slow walking speeds. For most cases, these deviations only lead to a small decline in performance. As already mentioned, it is assumed that all named limitations can be removed with suitable configuration and training data. For the detection of events, the use of neural networks is highly recommended. In terms of time difference, conventional methods can surpass the proposed algorithm. These methods are for offline detection only and differ greatly in experimental protocol. Despite some conventional methods being superior, the measured time difference of the proposed algorithm is very short. On average, events in healthy and clinical set are detected about $12\,ms$ earlier than indicated in the ground truth. This early detection benefits a processing in real-time. Also, time

differences of the proposed algorithm are more consistent than in existing methods as less deviation appears.
The last requirement was an automatic adaptation to the individual gait behavior of the observed subject to achieve the highest possible detection rate of gait events. This requirement has been met only partially. The individual LSTM RNNs are sufficiently generalized that an adaptation to the observed subject is not necessary for the investigated data set. Likewise, the proposed algorithm does not contain any parameters that an operator would have to tune for initialization or during run time. The aspect of adaptation is only taken up in post-processing for the detection of false positive events via self-adapting parameters. Post-processing works in itself well enough to get a good performance, thus, no further adaptive parameters were considered necessary. By contrast, in phase estimation via regression only a moderate performance was achieved. As already stated, the performance could be increased by a changed configuration or by change in method. A considered approach is a replication of the method of Zheng et al. [95] and Yan et al. [92]. In this approach, the gait phase of a stride would be predicted as an adaptive periodic model. Using detected events as landmarks, the model would be updated with an error signal at each detection, thus adapting to the individual gait pattern of the subject. In contrast to the approaches of Zheng et al. [95] and Yan et al. [92], this would be a self-adapting model an no parameter tuning through an operator would be necessary. It is assumed, that this individualized adaptation will yield a better performance for estimation of gait phases than an optimized regression via neural networks.

# 6 Conclusion

In order to improve the determination of gait phases in gait analysis, this thesis aimed to make a further contribution as investigating the suitability of long short term memory recurrent neural networks. An algorithm for the detection of gait events and the regression of percentual gait phases in real-time applications of optical marker-based systems was developed for this cause. The proposed algorithm consists of a group of several LSTM RNNs, each of which concentrates on one case of application on the corresponding body side. The algorithm is easy to integrate into existing applications, offers a high recognition rate of gait events with a considerable small delay and contains a self-adapting post-processing to further improve event detection rate in observed individuals.
In comparison with existing methods, LSTM RNNs achieve remarkable results in event detection for selected non-disabled subjects and individuals with gait pathologies regarding incomplete spinal cord injury and stroke survivors. In conventional, heuristic approaches, performance decreases, when gait data substantially deviates from normative values. In addition, most approaches only use a relatively small number of cases from a selected group of patients, which hinders a generalization. In this case, LSTM RNNs have a better generalization over the investigated subjects and surpass several conventional methods. However, the investigated sample size of the clinical consists also of only 23 subjects with two different pathologies. In the considered data set, the proposed algorithm covers 97.05% of all events in a scope of $50\,ms$. 98.88% of events are covered in a limitless scope. On average, the mean time deviation is $-11.83 \pm 6.01\,ms$ for detected *Initial Contact* and $-11.41 \pm 8.02\,ms$ for detected *Final Contact*. Although the proposed algorithm is superior to most of consulted methods, it is assumed, that the chosen configuration for the LSTM RNNs may not be optimal. This offers further room for improvement. In addition, the self-developed algorithm of the *Spinal Cord Injury Center* is surpassed. Performance of the self-developed algorithm is not up to existing methods and a further use is not recommended. The proposed algorithm is suited as a replacement for the self-developed algorithm. The results of this thesis are intended as a lead for further research to find more suitable configurations and architectures of LSTM RNNs. It is assumed, that the use of neural networks is indispensable for considerably improving the current state-of-the-art in event detection.
Regarding the estimation of gait phases, results have shown, that LSTM RNNs are less suitable for a regression of gait phases. The mean absolute error to the ground truth for regression of gait phases is $7.25 \pm 1.45\%$. For the estimation of gait phases via detected events, the mean absolute error is $1.95 \pm 1.10\%$. Only moderate performance is achieved and the approach of regression shows extensive inferiority to existing methods. For a use of the proposed algorithm, it is therefore recommended to estimate gait phases out of time distances between detected events. The performance of this approach is in line with the performance of existing methods for estimation of gait phases. Although it is possible to achieve a better performance with a suitable configuration, and thus superiority to existing methods, it is assumed, that the approach of regression does not offer the optimal solution. Results of existing methods [92, 95] have shown, that better performance can be achieved, if the stride is considered as a periodic signal rather a interdependent signal,

as is the case with LSTM RNNs. For future works for the estimation of gait phases, it is recommended to orient by these methods. However, these approaches only consider the percentual gait phase as a value between 0 and 1. Other options for the depiction of gait phases exists, but were not within the scope of this thesis.
Furthermore, this thesis aimed to determine whether a more sophisticated, data-driven processing guarantees a better performance over heuristic approaches and whether a direct calculation or a computation via detected events is more suitable for the estimation of gait phases. It has been revealed that conventional methods achieve similar performance with simpler processing. However, conventional methods are often used only for offline processing and do not generalize well between healthy subjects and individuals with gait pathologies. A data-driven processing through neural networks made it possible for the proposed algorithm to be flexible regarding observed subjects and to be used in real-time application. For estimation of gait phases, results have shown that direct calculation of gait phases yields better performance. Yet, a regression via LSTM RNNs is not appropriate.

## 6.1 Limitations

The main limitation of the proposed algorithm is also a general problem in gait analysis. The performance of the proposed algorithm decreases for slower walking speed, both in event detection and phase regression. Measurements M10146 and M10737 also revealed, that further movement of the foot prior *Initial Contact* as well as conducting *Initial Contact* with the toe or whole foot leads to a decrease in performance. The drop in performance for M10737 is much lower than in M10146, indicating that the proposed algorithm focuses more on the movement prior IC than on IC itself. As before, these patterns differ fundamentally from the learned gait pattern. It is expected, that severely impaired individuals of the trained gait pattern or individuals with gait pathologies beyond the trained scope will also constitute a limitation for the proposed algorithm. It is assumed, that all named limitations can be removed with an appropriate set of training data.

## 6.2 Future Work

For future works it is recommended to use a force plate for determining a ground truth, provided the elementary gait events *Initial Contact* and *Final Contact* shall be detected. Manual annotation of these events is considered to be too tedious and time-consuming. For a higher granularity of the gait phase with more gait events, manual annotation is, if no foot pressure insoles are available, still essential. Since results of manual annotation indicate a connection between rater agreement and walking speed of observed subject, this connection should be further investigated.
For the further development of the proposed algorithm, two ways have crystallized. On one hand, it was presented, that the used configuration for training the proposed algorithm is sub-optimal, both for event detection and phase regression. Therefore, it is advisable to revise the feature selection and to evaluate the selected features using a second feature selection method. Regardless of the feature selection, the development process should be continued in order to test further training parameters and architectures to find a more suitable configuration. It is also advisable to include gait data of children and runners, as well as individuals with gait pathologies beyond the trained scope in the training set in order to be able to train the algorithm for more gait patterns and achieve an overall higher generalization. If these proposals are implemented, the performance

in terms of event detection detection could be further improved and the aforementioned limitations removed. It should also be investigated whether increasing the sample rate of data recording or removing the used data filter in the proposed algorithm will yield lower temporal differences to the ground truth regarding detection of gait events. Finally, an evaluation of the proposed algorithm regarding event detection needs to be conducted against an established method for real-time estimation of gait phases under real-time conditions.

In terms of gait phase regression, a change in configuration would also increase performance. However, as already stated, it is assumed, that approaches like Zheng et al. [95] and Yan et al. [92] are better suited for determining gait phases. For this purpose, it would be appropriate to remove the corresponding LSTM RNNs from the proposed algorithm and replace them with an self-adaptive model for prediction of periodic signals. It is assumed, that this approach will yield the best currently possible performance regarding estimation of gait phases in real-time.

# References


[1] Erik K. Antonsson and Robert W. Mann. “The frequency content of gait”. In: *Journal of Biomechanics* 18.1 (1985), pp. 39–47. ISSN: 0021-9290. DOI: 10.1016/0021-9290(85)90043-0.

[2] Marco Avvenuti et al. “Smart Shoe-Assisted Evaluation of Using a Single Trunk / Pocket-Worn Accelerometer to Detect Gait Phases”. In: *Sensors* 18 (2018), p. 3811. ISSN: 1424-8220. DOI: 10.3390/s18113811.

[3] Filipe O. Barroso et al. “Combining muscle synergies and biomechanical analysis to assess gait in stroke patients”. In: *Journal of Biomechanics* 63 (2017), pp. 98–103. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2017.08.006.

[4] Imad A. Basheer and Maha Hajmeer. “Artificial neural networks: fundamentals, computing, design, and application”. In: *Journal of Microbiological Methods* 43.1 (2000), pp. 3–31. ISSN: 0167-7012. DOI: 10.1016/S0167-7012(00)00201-3.

[5] *Basic sciences.* Accessed: 2019-10-24. URL: https://musculoskeletalkey.com/basic-sciences/#f0010.

[6] Leonard E. Baum and Ted Petrie. “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”. In: *The Annals of Mathematical Statistics* 37.6 (1966), pp. 1554–1563. ISSN: 0003-4851. DOI: 10.2307/2238772.

[7] A. Behboodi et al. “Seven phases of gait detected in real-time using shank attached gyroscopes”. In: (2015), pp. 5529–5532. DOI: 10.1109/EMBC.2015.7319644.

[8] Richard Bellman. *Dynamic Programming.* 1st ed. Princeton University Press, 1957. ISBN: 9780691079516.

[9] Tor Jan Derek Berstad et al. “Tradeoffs Using Binary and Multiclass Neural Network Classification for Medical Multidisease Detection”. In: *2018 IEEE International Symposium on Multimedia (ISM)* (2018), pp. 1–8. DOI: 10.1109/ISM.2018.00009.

[10] Arturo Bertomeu-Motos. “Biomechanics of human walking and stability descriptive parameters”. In: *Revista Doctorado UMH. Vol* 1.1 (2015), pp. 1–6. ISSN: 2530-7320. DOI: 10.21134/doctumh.v1i1.880.

[11] Christopher M. Bishop. *Pattern recognition and machine learning.* 1st ed. Springer, 2006. ISBN: 1493938436.

[12] J. Martin Bland and Douglas G. Altman. “Measuring agreement in method comparison studies”. In: *Statistical Methods in Medical Research* 8.2 (1999), pp. 135–160. DOI: 10.1177/096228029900800204.

[13] J. Martin Bland and Douglas G. Altman. “Statistical methods for assessing agreement between two methods of clinical measurement”. In: *The Lancet* 327.8476 (1986), pp. 307–310. DOI: 10.1016/S0140-6736(86)90837-8.

[14] Sarah Boslaugh. *Statistics in a Nutshell.* 2nd ed. O’Reilly and Associates, 2012. ISBN: 1449316824.

[15] Salah Bouktif et al. “Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †”. In: *Energies* 11.7 (2018), p. 1636. ISSN: 1996-1073. DOI: 10.3390/en11071636.

[16] Gabriele Bovi et al. "A multiple-task gait analysis approach: Kinematic, kinetic and EMG reference data for healthy young and adult subjects". In: *Gait and Posture* 33.1 (2011), pp. 6–13. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2010.08.009.

[17] Dustin A. Bruening and Sarah Trager Ridge. "Automated event detection algorithms in pathological gait". In: *Gait and Posture* 39.1 (2014), pp. 472–477. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2013.08.023.

[18] O Bunke and B Droge. "Estimators of the mean squared error of prediction in linear regression". In: *Technometrics* 26.2 (1984), pp. 144–155. DOI: 10.1080/00401706.1984.10487940.

[19] Stephen Butterworth. "On the theory of filter amplifiers". In: *Wireless Engineer* 7.6 (1930), pp. 536–541. URL: https://www.changpuak.ch/electronics/downloads/On_the_Theory_of_Filter_Amplifiers.pdf.

[20] Junyoung Chung et al. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *CoRR* abs/1412.3555 (2014), pp. 1–7. URL: https://arxiv.org/abs/1412.3555.

[21] Zhicheng Cui, Wenlin Chen, and Yixin Chen. "Multi-Scale Convolutional Neural Networks for Time Series Classification". In: *CoRR* abs/1603.06995 (2016), pp. 1–10. URL: http://arxiv.org/abs/1603.06995.

[22] M. Dash and H. Liu. "Feature selection for classification". In: *Intelligent Data Analysis* 1.1 (1997), pp. 131–156. ISSN: 1088-467X. DOI: 10.1016/S1088-467X(97)00008-5.

[23] Eric Desailly et al. "Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait". In: *Gait and Posture* 29.1 (2009), pp. 76–80. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2008.06.009.

[24] Moataz Eltoukhy et al. "Improved kinect-based spatiotemporal and kinematic treadmill gait assessment". In: *Gait and Posture* 51 (2017), pp. 77–83. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2016.10.001.

[25] Rebecca E. Fellin et al. "Comparison of methods for kinematic identification of footstrike and toe-off during overground and treadmill running". In: *Journal of Science and Medicine in Sport* 13.6 (2010), pp. 646–650. ISSN: 1440-2440. DOI: 10.1016/j.jsams.2010.03.006.

[26] Deanna J. Fish and Jean-Paul Nielsen. "Clinical Assessment of Human Gait". In: *Journal of Prosthetics and Orthotics* 5.2 (1993), pp. 39–84. URL: https://journals.lww.com/jpojournal/Fulltext/1993/04000/Clinical_Assessment_of_Human_Gait.5.aspx.

[27] Randal C. Foster et al. "Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure". In: *Preventive Medicine* 41.3 (2005), pp. 778–783. ISSN: 0091-7435. DOI: 10.1016/j.ypmed.2005.07.006.

[28] Cardillo G. *Cohen's kappa: compute the Cohen's kappa ratio on a 2x2 matrix.* Accessed: 2019-07-02. URL: http://www.mathworks.com/matlabcentral/fileexchange/15365.

[29] Cardillo G. *Fleiss'es kappa: compute the Fleiss'es kappa for multiple raters.* Accessed: 2019-07-24. URL: http://www.mathworks.com/matlabcentral/fileexchange/15426.

[30] James R. Gage, Peter A. Deluca, and Thomas S. Renshaw. "Gait Analysis: Principles and Applications". In: *The Journal of Bone and Joint Surgery* 77-A.10 (1995), pp. 1607–1623. URL: https://journals.lww.com/jbjsjournal/Fulltext/1995/10000/Gait_Analysis__Principles_and_Applications.17.aspx.

[31] F. A. Gers, J. Schmidhuber, and F. Cummins. "Learning to forget: continual prediction with LSTM". In: 2 (1999), pp. 850–855. DOI: 10.1049/cp:19991218.

[32] Felix Gers. "Long short-term memory in recurrent neural networks". PhD thesis. Welfengarten 1, 30167 Hannover, Germany: Gottfried Wilhelm Leibniz Universität Hannover, 2001.

[33] Salim Ghoussayni et al. "Assessment and validation of a simple automated method for the detection of gait events and intervals". In: *Gait and Posture* 20.3 (2004), pp. 266–272. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2003.10.001.

[34] Richard H. R. Hahnloser et al. "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit". In: *Nature* 405 (2000), 947–951. ISSN: 1476-4687. DOI: 10.1038/35016072.

[35] Andrew H. Hansen, Dudley S. Childress, and Margrit R. Meier. "A simple method for determination of gait events". In: *Journal of Biomechanics* 35.1 (2002), pp. 135–138. ISSN: 0021-9290. DOI: 10.1016/S0021-9290(01)00174-9.

[36] Jeffrey M. Hausdorff. "Gait variability: methods, modeling and meaning". In: *Journal of NeuroEngineering and Rehabilitation* 2.19 (2005), pp. 1–9. ISSN: 1743-0003. DOI: 10.1186/1743-0003-2-19.

[37] *hawk-i Digital RealTime System*. Accessed: 2019-10-17. URL: http://www.hoskinscientifique.com/uploadpdf/Instrumentation/Motion%20Analysis/hoskin_hawki_4bacfe3ceb134.pdf.

[38] Brad D. Hendershot, Caitlin E. Mahon, and Alison L. Pruziner. "A comparison of kinematic-based gait event detection methods in a self-paced treadmill application". In: *Journal of Biomechanics* 49.16 (2016), pp. 4146–4149. DOI: 10.1016/j.jbiomech.2016.10.046.

[39] Tin Kam Ho. "Random decision forests". In: *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1.1 (1995), pp. 278–282. DOI: 10.1109/ICDAR.1995.598994.

[40] Mohammad Hossin and MN Sulaiman. "A review on evaluation metrics for data classification evaluations". In: *International Journal of Data Mining and Knowledge Management Process* 5.2 (2015), pp. 1–11. DOI: 10.5121/ijdkp.2015.5201.

[41] Alan Hreljac and Robert N Marshall. "Algorithms to determine event timing during normal walking using kinematic data". In: *Journal of Biomechanics* 33.6 (2000), pp. 783–786. ISSN: 0021-9290. DOI: 10.1016/S0021-9290(00)00014-2.

[42] Alan Hreljac and Nicholas Stergiou. "Phase determination during normal running using kinematic data". In: *Medical and Biological Engineering and Computing* 38.5 (2000), pp. 503–506. ISSN: 0140-0118. DOI: 10.1007/bf02345744.

[43] B-J Hsue et al. "Gait timing event determination using kinematic data for the toe walking children wtih cerebral palsy". In: *Journal of Biomechanics* 40 (2007), S529. ISSN: 0021-9290. DOI: 10.1016/S0021-9290(07)70519-5.

[44] Yanbo Huang. "Advances in Artificial Neural Networks – Methodological Development and Application". In: *Algorithms* 2.3 (2009), 973–1007. ISSN: 1999-4893. DOI: 10.3390/algor2030973.

[45] Verne Thompson Inman, Henry James Ralston, and Frank Todd. *Human walking*. Williams & Wilkins, 1981.

[46] Anil K. Jain, Jianchang Mao, and K. Moidin Mohiuddin. "Artificial Neural Networks: A Tutorial". In: *Computer* 29.3 (1996), pp. 31–44. ISSN: 0018-9162. DOI: 10.1109/2.485891.

[47] Xianta Jiang et al. "A Wearable Gait Phase Detection System Based on Force Myography Techniques". In: *Sensors* 18.4 (2018), p. 1279. ISSN: 1424-8220. DOI: 10.3390/s18041279.

[48] M. P. Kadaba, H. K. Ramakrishnan, and M. E. Wootten. "Measurement of lower extremity kinematics during level walking". In: *Journal of Orthopaedic Research* 8.3 (1990), pp. 383–392. DOI: 10.1002/jor.1100080310.

[49] F. Karim et al. “LSTM Fully Convolutional Networks for Time Series Classification”. In: *IEEE Access* 6 (2018), pp. 1662–1669. DOI: 10.1109/ACCESS.2017.2779939.

[50] T. Karčnik. “Using motion analysis data for foot-floor contact detection”. In: *Medical and Biological Engineering and Computing* 41 (2003), pp. 509–512. ISSN: 0140-0118. DOI: 10.1007/BF02345310.

[51] Łukasz Kidziński, Scott Delp, and Michael Schwartz. “Automatic real-time gait event detection in children using deep neural networks”. In: *PLOS ONE* 14 (2019), pp. 1–11. DOI: 10.1371/journal.pone.0211466.

[52] Terry K. Koo and Mae Y. Li. “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research”. In: *Journal of Chiropractic Medicine* 15.2 (2016), pp. 155–163. DOI: 10.1016/j.jcm.2016.02.012.

[53] D. Kotiadis, H.J. Hermens, and P.H. Veltink. “Inertial Gait Phase Detection for control of a drop foot stimulator Inertial sensing for gait phase detection”. In: *Medical Engineering and Physics* 32.4 (2010), pp. 287–297. DOI: 10.1016/j.medengphy.2009.10.014.

[54] Robert Kwiecien, Annette Kopp-Schneider, and Maria Blettner. *Konkordanzanalyse.* 2011. DOI: 10.3238/arztebl.2011.0515.

[55] J. Richard Landis and Gary G. Koch. “An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers”. In: *Biometrics* 33.2 (1977), pp. 363–374. ISSN: 1541-0420. DOI: 10.2307/2529786.

[56] Jessica Leitch et al. “Identifying gait events without a force plate during running: A comparison of methods”. In: *Gait and Posture* 33.1 (2011), pp. 130–132. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2010.06.009.

[57] Andrea Mannini and Angelo Maria Sabatini. “Gait phase detection and discrimination between walking–jogging activities using hidden Markov models applied to foot motion data from a gyroscope”. In: *Gait and Posture* 36.4 (2012), pp. 657–661. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2012.06.017.

[58] Benoit Mariani et al. “Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors”. In: *Gait and Posture* 37.2 (2013), pp. 229–234. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2012.07.012.

[59] MathWorks. *Deep Learning Toolbox.* Accessed: 2019-01-15. URL: https://de.mathworks.com/products/deep-learning.html.

[60] MathWorks. *fitcensemble.* Accessed: 2019-01-15. URL: https://de.mathworks.com/help/stats/fitcensemble.html.

[61] MathWorks. *fitrensemble.* Accessed: 2019-01-15. URL: https://de.mathworks.com/help/stats/fitrensemble.html.

[62] MathWorks. *oobPermutedPredictorImportance.* Accessed: 2019-01-15. URL: https://de.mathworks.com/help/stats/regressionbaggedensemble.oobpermutedpredictorimportance.html.

[63] MathWorks. *predictorImportance.* Accessed: 2019-01-15. URL: https://de.mathworks.com/help/stats/compactregressiontree.predictorimportance.html.

[64] MathWorks. *templateTree.* Accessed: 2019-01-15. URL: https://de.mathworks.com/help/stats/templatetree.html#d117e848665.

[65] Kenneth O. McGraw and S. P. Wong. “Forming inferences about some intraclass correlations coefficients”. In: *Psychological Methods* 1.1 (1996), pp. 30–46. DOI: 10.1037/1082-989X.1.4.390.

[66] J Mickelborough et al. “Validity and reliability of a kinematic protocol for determining foot contact events”. In: *Gait and Posture* 11.1 (2000), pp. 32–37. ISSN: 0966-6362. DOI: 10.1016/S0966-6362(99)00050-8.

[67] Adam Miller. "Gait event detection using a multilayer neural network". In: *Gait and Posture* 29.4 (2009), pp. 542–545. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2008.12.003.

[68] MotionAnalysis. *OrthoTrak 6.5 Gait Analysis Software Reference Manual*. Motion Analysis Corporation. 3617 Westwind Boulevard, Santa Rosa, CA 95403 USA, 2007.

[69] Ciara M. O'Connor et al. "Automatic detection of gait events using kinematic data". In: *Gait and Posture* 25.3 (2007), pp. 469–474. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2006.05.016.

[70] Sean T. Osis et al. "Predicting timing of foot strike during running, independent of striking technique, using principal component analysis of joint angles". In: *Journal of Biomechanics* 47.11 (2014), pp. 2786–2789. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2014.06.009.

[71] I. P. I. Pappas et al. "A reliable gait phase detection system". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9.2 (2001), pp. 113–125. DOI: 10.1109/7333.928571.

[72] Y. Qi et al. "Assessment of Foot Trajectory for Human Gait Phase Detection Using Wireless Ultrasonic Sensor Network". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24.1 (2016), pp. 88–97. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2015.2409123.

[73] Ines A. Kramers de Quervain, Edgar Stüssi, and Alex Stacoff. "Ganganalyse beim Gehen und Laufen". In: *Schweizerische Zeitschrift für «Sportmedizin und Sporttraumatologie»* 56.2 (2008), pp. 35–42. URL: https://www.researchgate.net/profile/Ines_Kramers-De_Quervain/publication/267679143_35_Ganganalyse_beim_Gehen_und_Laufen/links/552531810cf2b123c5179949.pdf.

[74] Arash Salarian. *Intraclass Correlation Coefficient (ICC)*. Accessed: 2019-07-26. URL: https://de.mathworks.com/matlabcentral/fileexchange/22099-intraclass-correlation-coefficient-icc.

[75] Daniel Schließmann et al. "Trainer in a pocket - Proof-of-concept of mobile, real-time, foot kinematics feedback for gait pattern normalization in individuals after stroke, incomplete spinal cord injury and elderly patients". In: *Journal of Neuro-Engineering and Rehabilitation* 15.44 (2018), pp. 1–15. DOI: 10.1186/s12984-018-0389-4.

[76] C. M. Senanayake and S. M. N. A. Senanayake. "Computational Intelligent Gait-Phase Detection System to Identify Pathological Gait". In: *IEEE Transactions on Information Technology in Biomedicine* 14.5 (2010), pp. 1173–1179. DOI: 10.1109/TITB.2010.2058813.

[77] Patrick E Shrout and Joseph L Fleiss. "Intraclass correlations: uses in assessing rater reliability". In: *Psychological bulletin* 86.2 (1979), pp. 420–428. DOI: 10.1037/0033-2909.86.2.420.

[78] Lisa A. Simpson et al. "Capturing step counts at slow walking speeds in older adults: Comparison of ankle and waist placement of measuring device". In: *Journal of Rehabilitation Medicine* 47.9 (2015), 830–835. ISSN: 1650-1977. DOI: 10.2340/16501977-1993.

[79] M. M. Skelly and H. J. Chizeck. "Real-time gait event detection for paraplegic FES walking". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9.1 (2001), pp. 59–68. DOI: 10.1109/7333.918277.

[80] Denis Smirnov and Engelbert Mephu Nguifo. "Time Series Classification with RecurrentNeural Networks". In: (2018), pp. 1–8. URL: https://project.inria.fr/aaldt18/files/2018/08/oral.pdf.

[81] D. Stathakis. "How many hidden layers and nodes?" In: *International Journal of Remote Sensing* 30.8 (2009), pp. 2133–2147. DOI: 10.1080/01431160802549278.

[82] Juri Taborri et al. “Gait Partitioning Methods: A Systematic Review”. In: *Sensors* 16.66 (2016), pp. 1–20. ISSN: 1424-8220. DOI: 10.3390/s16010066.

[83] Diana Trojaniello et al. “Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait”. In: *Journal of NeuroEngineering and Rehabilitation* 11.152 (2014), pp. 1–12. ISSN: 1743-0003. DOI: 10.1186/1743-0003-11-152.

[84] Susanne M. van der Veen et al. “Validation of gait event detection by centre of pressure during target stepping in healthy and paretic gait”. In: *Journal of Biomechanics* 79 (2018), pp. 218–222. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2018.07.039.

[85] A. Viterbi. “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”. In: *IEEE Transactions on Information Theory* 13.2 (1967), pp. 260–269. ISSN: 1557-9654. DOI: 10.1109/TIT.1967.1054010.

[86] Huong Thi Thu Vu et al. “ED-FNN: A New Deep Learning Algorithm to Detect Percentage of the Gait Cycle for Powered Prostheses”. In: *Sensors* 18.7 (2018), p. 2389. DOI: 10.3390/s18072389.

[87] Z. Wang, W. Yan, and T. Oates. “Time series classification from scratch with deep neural networks: A strong baseline”. In: (2017), pp. 1578–1585. DOI: 10.1109/IJCNN.2017.7966039.

[88] Jamie L. S. Waugh. “Online Learning of Gait Models”. PhD thesis. 200 University Ave W, Waterloo ON N2L 3G1, Canada: University of Waterloo, 2018.

[89] Michael W. Whittle. “Clinical gait analysis: A review”. In: *Human Movement Science* 15.3 (1996), pp. 369–387. ISSN: 0167-9457. DOI: 10.1016/0167-9457(96)00006-1.

[90] R. Williamson and B. J. Andrews. “Gait event detection for FES using accelerometers and supervised machine learning”. In: *IEEE Transactions on Rehabilitation Engineering* 8.3 (2000), pp. 312–319. DOI: 10.1109/86.867873.

[91] Cort J Willmott and Kenji Matsuura. “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance”. In: *Climate research* 30.1 (2005), pp. 79–82. ISSN: 1616-1572. DOI: 10.3354/cr030079.

[92] Tingfang Yan et al. “An oscillator-based smooth real-time estimate of gait phase for wearable robotics”. In: *Autonomous Robots* 41.3 (2017), pp. 759–774. ISSN: 1573-7527. DOI: 10.1007/s10514-016-9566-0.

[93] Lie Yu et al. “Adaptive method for real-time gait phase detection based on ground contact forces”. In: *Gait and Posture* 41.1 (2015), pp. 269–275. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2014.10.019.

[94] J.A. Zeni, J.G. Richards, and J.S. Higginson. “Two simple methods for determining gait events during treadmill and overground walking using kinematic data”. In: *Gait and Posture* 27.4 (2008), pp. 710–714. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2007.07.007.

[95] E. Zheng et al. “Gait Phase Estimation Based on Noncontact Capacitive Sensing and Adaptive Oscillators”. In: *IEEE Transactions on Biomedical Engineering* 64.10 (2017), pp. 2419–2430. ISSN: 0018-9294. DOI: 10.1109/TBME.2017.2672720.

[96] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. “A survey on unsupervised outlier detection in high-dimensional numerical data”. In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5.5 (2012), pp. 363–387. ISSN: 1932-1872. DOI: 10.1002/sam.11161.

# A Training Parameter and Architecture of proposed Networks

### Network for Detection of Gait Events

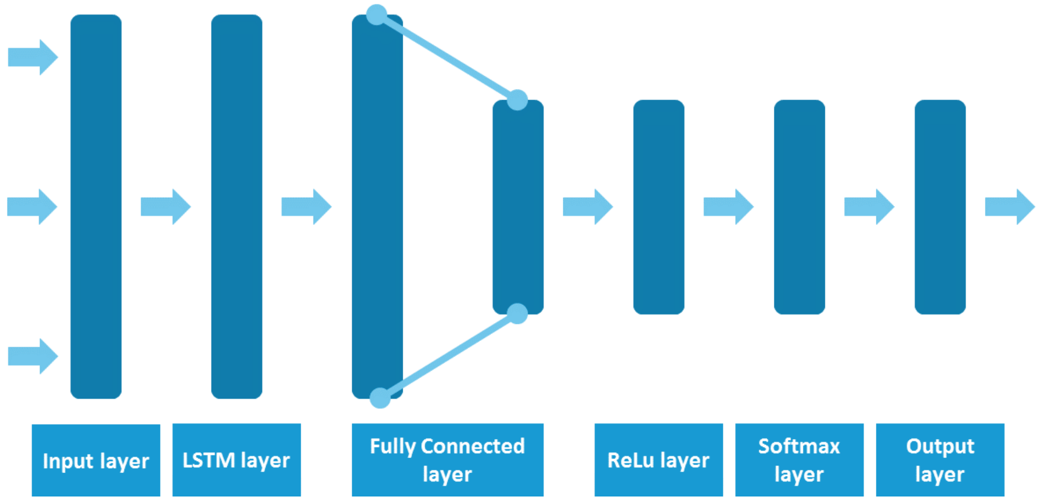


FIGURE A.1: Pipeline of data processing of LSTM RNN cell for detection of gait events in the proposed algorithm

Figure A.1 shows the pipeline of data processing within a LSTM RNN for detection of gait events. The processing is as follows:

1. Network input is handled by a sequence input layer. The input consists of a sample of the currently considered frame in eight dimensions. These dimensions correspond to the required features, which are displayed in Table 3.2.

2. The input sample is processed within a LSTM RNN cell. For detection of gait events, the cell contains 256 hidden units. Cell output is only the last value of finished processing.

3. Using a fully connected layer, the output of the LSTM RNN cell is mapped to an array with three values. These values represent the three possible event classes, that the network can detect. The event classes are *Initial Contact*, *Final Contact* and *no event*.

4. The three values from the fully connected layer are adjusted with a so-called *Rectified Linear Unit layer* [34]. In this layer, negative array values are overwritten with the value 0, while positive values remain unchanged.

5. Within the softmax layer [11], the three values are scaled to a range between 0 and 1. Through the preparation of the ReLu layer, these three values can now be interpreted as probabilities.
6. In the output classification layer, the class with the highest probability is selected as the output of the network

For an exact reproduction of the network, Table A.1 shows the used configuration for the training of the network. Unspecified parameters were kept with their default settings.

Table A.1: Training parameter of LSTM RNN cell for detection of gait events in the proposed algorithm

| **Parameter** | **Value** |
|---:|:---:|
| Solver | adam |
| MaxEpochs | 60 |
| Shuffle | never |
| MiniBatchSize | 128 |
| Epsilon | 1e-8 |
| GradientThreshold | 0.25 |

**Network for Regression of Gait Phases**

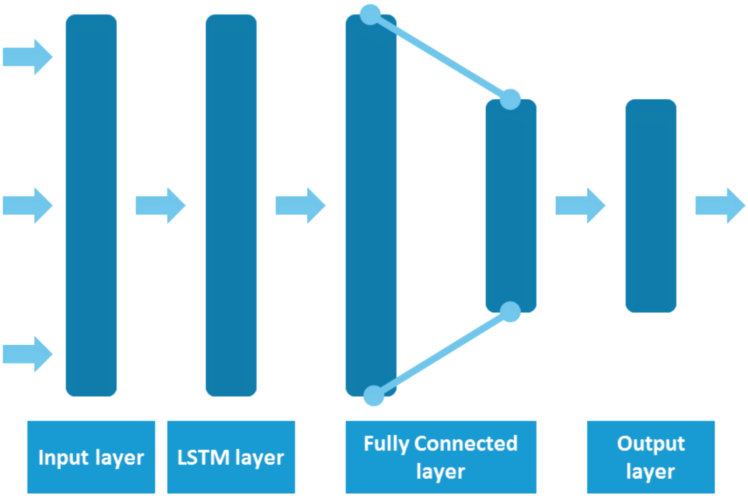


Figure A.2: Pipeline of data processing of LSTM RNN cell for regression of gait phases in the proposed algorithm

Figure A.2 shows the pipeline of data processing within a LSTM RNN for regression of gait phases. The processing is as follows:

1. The input of the LSTM RNN is handled by a sequence input layer. The input consists of a sample of the currently considered frame in 15 dimensions, which correspond to the required features. The features are displayed in Table 3.3.

2. All input is processed within a LSTM RNN cell with 512 hidden units. The output of the cell is the last value of finished processing.
3. Using a fully connected layer, the output of the LSTM RNN cell is mapped to a single value. This value corresponds to the phase in the gait cycle
4. Through an output regression layer, the gait phase is returned as the output of the network.

For an exact reproduction of the network, Table A.2 shows the used configuration for the training of the network. Unspecified parameters were kept with their default settings.

TABLE A.2: Training parameter of LSTM RNN cell for regression of gait phases in the proposed algorithm

| **Parameter** | **Value** |
|---:|:---:|
| Solver | adam |
| MaxEpochs | 30 |
| Shuffle | never |
| MiniBatchSize | 128 |
| Epsilon | 1e-8 |
| GradientThreshold | Inf |